Genetic and Statistical Analyses of Strong Selection on Polygenic Traits: What, Me Normal?

Michael Turelli* and N. H. Barton†

*Section of Evolution and Ecology and Center for Population Biology, University of California, Davis, California 95616 and
†Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom

ABSTRACT

We develop a general population genetic framework for analyzing selection on many loci, and apply it to strong truncation and disruptive selection on an additive polygenic trait. We first present statistical methods for analyzing the infinitesimal model, in which offspring breeding values are normally distributed around the mean of the parents, with fixed variance. These show that the usual assumption of a Gaussian distribution of breeding values in the population gives remarkably accurate predictions for the mean and the variance, even when disruptive selection generates substantial deviations from normality. We then set out a general genetic analysis of selection and recombination. The population is represented by multilocus cumulants describing the distribution of haploid genotypes, and selection is described by the relation between mean fitness and these cumulants. We provide exact recursions in terms of generating functions for the effects of selection on non-central moments. The effects of recombination are simply calculated as a weighted sum over all the permutations produced by meiosis. Finally, the new cumulants that describe the next generation are computed from the non-central moments. Although this scheme is applied here in detail only to selection on an additive trait, it is quite general. For arbitrary epistasis and linkage, we describe a consistent infinitesimal limit in which the short-term selection response is dominated by infinitesimal allele frequency changes and linkage disequilibria. Numerical multilocus results show that the standard Gaussian approximation gives accurate predictions for the dynamics of the mean and genetic variance in this limit. Even with intense truncation selection, linkage disequilibria of order three and higher never cause much deviation from normality. Thus, the empirical deviations frequently found between predicted and observed responses to artificial selection are not caused by linkage-disequilibrium-induced departures from normality. Disruptive selection can generate substantial four-way disequilibria, and hence kurtosis; but even then, the Gaussian assumption predicts the variance accurately. In contrast to the apparent simplicity of the infinitesimal limit, data suggest that changes in genetic variance after 10 or more generations of selection are likely to be dominated by allele frequency dynamics that depend on genetic details.

M OST analyses of selection on polygenic traits assume that the joint distribution of phenotypes and of breeding values is approximately Gaussian, once an appropriate scale of measurement is chosen. This ensures that the average phenotype of the offspring depends linearly on the phenotypes of the two parents and implies the standard equation for the response of the mean to selection:

$$\Delta \bar{Z} = h^2 \Delta_s \bar{Z}, \qquad (1)$$

(*i.e.*, $R = h^2 S$). Here $\Delta \bar{Z}$ is the selection response (R, the between-generation change in the mean); $h^2 = V_A / V_P$ is the (narrow sense) heritability, the fraction of the phenotypic variance attributable to additive genetic effects; and $\Delta_s \bar{Z}$ is the selection differential [S, the within-generation change in the mean caused by selection, *cf.* BULMER (1980, Ch. 9)].

FISHER (1918) reconciled the Gaussian statistical description of the inheritance of quantitative traits with Mendelian genetics by assuming a very large number of unlinked loci, each with small additive effects. This gives the "infinitesimal model," in which each cross produces offspring whose phenotypes are normally distributed around the mean of the parents, with a fixed variance due to independent segregation at many loci. It does not, however, justify the multivariate Gaussian assumption that leads to Equation 1, because selection will generally distort the distribution of breeding values away from a Gaussian (BULMER 1980, Ch. 9). TURELLI and BARTON (1990) argued that such distortions, which arise even under weak selection through the generation of third and higher order linkage disequilibria, could in principle be substantial when selection is strong. We also began to develop a multilocus genetic approach to polygenic selection that includes both allele frequency changes and linkage disequilibria. The infinitesimal model can predict only departures from normality caused by linkage disequilibria and not those caused by epistasis. It also cannot describe the changes in within-family genetic variance caused by the allele frequency

changes that must occur with selection on a finite number of loci. This paper develops general statistical and multilocus population genetic methods for understanding the effects of selection on polygenic traits without assuming normality, and applies these to additive genetic models to determine when and why the standard Gaussian methods are accurate. Our basic conclusion, which could not have been foreseen from our weak selection analysis, is that even intense truncation selection on an additive polygenic trait is not likely to produce significant departures from (1) caused by higher order disequilibria.

We begin by setting out two "statistical" methods for deriving numerical results from the infinitesimal model: one based on iterating the distribution of breeding values, and the other based on the cumulants of this distribution (*cf.* ZENG 1987). We then generalize our previous genetic analyses (TURELLI and BARTON 1990; BARTON and TURELLI 1991) to provide exact equations for the dynamics of the means, variances, and higher-order cumulants under selection on an additive polygenic trait. We set out a general algorithm that describes the dynamics of cumulants of arbitrary order. We give explicit equations for changes in the first four cumulants (mean, variance, skewness and kurtosis) that involve selection coefficients up to fourth-order (defined below) and cumulants (and disequilibria) up to eighth order. These are checked by comparison with the infinitesimal model, and with exact numerical iterations of gamete frequencies for up to 100 loci, in which fitness is a polynomial function of the phenotypic (or genotypic) value. We use these results to approximate the consequences of two forms of phenotypic selection: truncation selection, because of its practical importance to plant and animal breeders, and disruptive selection, both as a model of speciation and because of its capacity to produce large departures from normality. Our analytical approximations for these non-polynomial selection schemes are checked against deterministic multilocus numerical calculations. We also extend the infinitesimal limit to allow for non-additive gene action.

The goal of our genetic analyses is to develop general and tractable methods for understanding multilocus selection. We introduce three innovations. First, we combine our general analysis of multilocus selection (BARTON and TURELLI 1991) with a description of selection in terms of gradients in mean fitness (BARTON and TURELLI 1987; TURELLI and BARTON 1990). Second, we give the equations in terms of cumulants rather than moments [a multilocus generalization of BÜRGER (1991)]. Finally, we have automated the intimidating algebra using the *Mathematica* symbolic computation language [WOLFRAM (1991), notebooks for the Macintosh are available on request]. We show that there is a class of models, which includes truncation selection on an additive polygenic trait, for which the distribution of breeding values remains close to Gaussian even when

selection is intense. In these cases, dynamics can be accurately predicted in terms of the gradients in mean fitness with respect to the population mean and variance. The short-term effects of linkage disequilibria on genetic variance are described by BULMER's (1971, 1980) extension of the infinitesimal model, while allele frequencies change more slowly, and depend on the detailed distribution of effects of each locus.

We concentrate on the simplest case of a trait determined by the sum of effects of alleles at many loci, plus a normally distributed environmental component. Our analysis of strong selection with simple genetics is complementary to NAGYLAKI's (1993) analysis of weak selection with more complex genetics and BÜRGER's (1993) and ZHIVOTOVSKY and GAVRILETS' (1992) analyses of exponential and quadratic selection. We assume discrete generations, diploidy, autosomal inheritance, random mating and viability selection [see BARTON and TURELLI (1991, Fig. 1)]. However, our methods can be applied to more general issues in population genetics theory. They readily extend to arbitrary patterns of natural and sexual selection (BARTON and TURELLI 1991). Our results suggest that with polygenic inheritance, higher order interactions can often be neglected, allowing multilocus systems to be understood in terms of allele frequencies and pairwise linkage disequilibria.

STATISTICAL ANALYSES:
A NON-GAUSSIAN INFINITESIMAL MODEL

The simplest model for the inheritance of quantitative traits assumes that within each family, the breeding values of sibs follow a normal distribution with fixed variance, and mean equal to the average of the breeding values of the two parents. This is known as the infinitesimal model, because it emerges when the trait is the sum of infinitesimal contributions from an infinite number of unlinked genes (BULMER 1980). The model further assumes that all genotypes experience independent and identically distributed environmental contributions that are normally distributed with mean 0 and variance V_E. Under the infinitesimal model, the variance within families is half the variance due to segregation at individual loci (denoted $V_{G,LE}$, the "genic variance," which is the genetic variance at linkage equilibrium). The distribution of breeding values in the population tends towards a normal distribution with variance $V_{G,LE}$. However, selection can generate linkage disequilibria that alter the genetic variance and distort the distribution away from normality.

Even with this simple model, understanding the effects of strong truncation selection is not trivial. Let $\Psi(g)$ denote the distribution (probability density function) of breeding values among zygotes before selection. (Table 1 provides a glossary of notation.) Selection just weights $\Psi(g)$ by $W(g)$, the fitness function for breeding values, giving $\Psi^*(g) = \Psi(g)W(g)/\bar{W}$, with $\bar{W} = \Psi(g)W(g)dg$. Reproduction (meiosis and random

TABLE 1

Glossary of repeatedly used notation

Symbol	Usage, (relevant equation in the text)
C_U^*	Non-central moment, the expectation of the product of contributions of loci in the set U: $E(x_U)$, (11b)
$C_{U,V}^*$	Non-central moment for diploids, involving maternally inherited alleles at the loci in U and paternally inherited alleles at the loci in V, (19a)
C_U	Central moment, the expectation of the product of deviations from the mean over loci in the set U, (55, APPENDIX A)
$f(\mathbf{x})$	Frequency of newly produced haploid products of meiosis with genotype $\mathbf{x}$; sometimes used as shorthand for $f(\mathbf{x}, \mathbf{x}^*)$, (14)
$f(\mathbf{x}, \mathbf{x}^*)$	Frequency of diploids with genes from the mother in state $\mathbf{x}$, from the father in state $\mathbf{x}^*$, see paragraph following (12c)
$\tilde{f}(\tilde{\mathbf{x}})$	Fourier transform of $f(\mathbf{x})$, (6b)
$\hat{f}((\hat{\mathbf{x}})$	Moment generating function of $f(\mathbf{x})$, (11)
h^2	Heritability, $V_A/V_P \equiv V_G/V_P$ for the additive model (20), (1).
$H_i(y)$	ith Hermite polynomial, (3)
i, j, k	Label individual loci
$I(p)$	"Selection intensity" under truncation selection with a fraction p of the population selected, (34)
K_i	ith cumulant of the distribution of breeding values, (8)
$K(\hat{x})$	Cumulant generating function, (12)
$\mathcal{L}_i$	Under the additive model, the selection gradient with respect to the ith cumulant of the distribution of breeding values or phenotypes (24, 26), note that $\mathcal{L}_j = \partial \ln(\bar{W})/\partial K_j = \mathcal{L}_{S,T}$ with $j = |S+T|$
$\mathcal{L}_{S,T}$	Selection gradient with respect to the cumulant $\kappa_{S,T}$, $\partial \ln(\bar{W})/\partial \kappa_{S,T}$, (19)
$L_{U,V}$	Selection gradient with respect to the central moment $C_{U,V}$, $\partial \log(\bar{W})/\partial C_{U,V}$, (APPENDIX A)
p	Under truncation selection, the fraction of the population selected, (34)
$r_{S,T}$	For disjoint sets S and T, this denotes the frequency of recombination events that combine alleles from one parent at loci in S with alleles from the other parent at loci in T, (13)
S, T, U, V	Label sets of loci: $e.g.$, $S = \{iij\}$
$U+V$	Concatenation of the elements in sets U and V, $e.g.$, if $U = \{ii\}$ and $V = \{ij\}$, $U+V = \{iiij\}$, (APPENDIX A)
$U-V$	Set obtained by deleting the elements of V from U, $e.g.$, if $U = \{iiijk\}$ and $V = \{ij\}$, $U-V = \{iik\}$, (APPENDIX A)
$|U|$	Number of elements in U: $e.g.$, $|\{iij\}| = 3$, (11b)
V_E	Environmental variance
V_G	Additive genetic variance: for the additive model, $V_G = 2 \sum_{ij} \kappa_{ij}$
$V_{G,LE}$	Genic variance: for the additive model, $V_{G,LE} = 2 \sum \kappa_{ii}$
V_P	Phenotypic variance: $V_P = V_G + V_E$
$W(\mathbf{X}, \mathbf{X}^*)$	Fitness (viability) of genotype $\mathbf{X}, \mathbf{X}^*$, (14a)
$\bar{W}$	Mean fitness, (2, 14a)
$\mathbf{X}$	A vector denoting a haploid genotype: $(X_1, X_2, \ldots, X_n)$, after (12c)
X_i	Variable indicating a haploid genotype at locus i; when used to describe events within a generation, it refers to a maternally derived gamete, after (12c) and (14)
X_U	Product of X_i over the set U: $\prod_{i \in u} X_i$, (11b)
X_i^*	A haploid genotype at locus i in a paternally derived gamete, after (12c)
$\bar{Z}$	Population mean, (1) and (20)
γ_3	Normalized skew of the distribution of breeding values, $E[(G - \bar{Z})^3]/V_G^{3/2}$
γ_4	Kurtosis of the distribution of breeding values, $E[(G - \bar{Z})^4]/V_G^2 - 3$
$\Delta C_{U,V}^*$	Change in the non-central moment $C_{U,V}^*$ between generations, (19a)
$\Delta \kappa_U$	Change in the cumulant κ_U between generations that would be observed if selection were so weak that products of the $\mathcal{L}_i$ could be ignored, (36–41)
κ_U	Multivariate cumulant, (11, 12), $e.g.$, κ_{ii} = variance at locus i
$\phi(y)$	Standard Gaussian density, $\exp(-y^2/2)/\sqrt{2\pi}$, (3)
$\Phi(x)$	Cumulative distribution of the standard normal, $\int_{-\infty}^x \phi(y)\,dy$,
$\Psi(g)$	Probability density function of breeding values among zygotes before selection, (2)

mating) then gives the distribution in the next generation as

$$\Psi^{**}(g) = \frac{1}{\sqrt{\pi V_{G,LE}}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi^*(x)\Psi^*(y)$$
$$\exp\left(-\frac{(g - (x+y)/2)^2}{V_{G,LE}}\right) dx\,dy.$$

(2)

[$cf.$ SLATKIN (1970) and KARLIN (1979)].

We will describe two approaches for calculating the net effect of selection and reproduction, without assuming that the distribution of breeding values, $\Psi(g)$, is Gaussian. The first, introduced abstractly by CORNISH and FISHER (1937) and applied most thoroughly by ZENG (1987), approximates the distribution by a Gram-Charlier expansion [STUART and ORD (1987, Ch. 6)]. Reproduction causes a simple change to the cumulants [BULMER (1980, pp. 148–149)], and hence in the

coefficients of the expansion. The second method is based on the fact that Equation 2 corresponds to a product of Fourier transforms. Because truncation selection does not cause much distortion from normality, the Gram-Charlier method works well for this case. It remains surprisingly accurate even when the distribution departs substantially from normality. We will illustrate this by comparing the Fourier transform and Gram-Charlier methods for strong disruptive selection, which can produce substantial deviations from normality even with free recombination and infinitely many loci.

Gram-Charlier approximation: Let G denote the breeding value, $\bar{Z}$ the population mean and V_G the additive genetic variance (which generally differs from $V_{G,LE}$ because of linkage disequilibrium). We approximate the distribution of standardized breeding values, $Y = (G - \bar{Z})/\sqrt{V_G}$, by

$$\Psi(y) = \phi(y)\left[1 + \sum_{i=3}^{imax} \frac{c_i}{i!} H_i(y)\right], \qquad (3)$$

where $\phi(y) = \exp(-y^2/2)/\sqrt{2\pi}$ is the standard Gaussian density and $H_i(y)$ is the ith Hermite polynomial. (When the same quantity is treated as both a random variable and a specific value, we denote the random variable by a capital letter and specific values by lower case.) The Hermite polynomials are defined by $d^i\phi(y)/dy^i = (-1)^i H_i(y)\phi(y)$; thus, $H_0(y) = 1$, $H_1(y) = y$, etc., with H_i a polynomial of order i. The coefficients c_3, c_4 and c_5 in (3) equal the cumulants K_3, K_4 and K_5 (discussed further below); the relation is slightly more complicated for the higher coefficients [see STUART and ORD (1987, Ch. 6) for a discussion of both the H_i and Equation 3].

Applying truncation selection so that only individuals with phenotype at least t units above the mean survive, the average fitness as a function of breeding value is

$$W(g) = \Phi\left(\frac{g - \bar{Z} - t}{\sqrt{V_E}}\right) \qquad (4a)$$

or

$$W(y) = \Phi\left((y - T)\sqrt{\frac{V_G}{V_E}}\right), \qquad (4b)$$

where T is the truncation point scaled relative to V_G (i.e., $T = t/\sqrt{V_G}$) and $\Phi(x) = \int_{-\infty}^{x} \phi(y)\,dy$ is the cumulative distribution of the standard normal. Equation 4 follows from the assumption that environmental deviations are Gaussian with mean 0 and variance V_E (non-Gaussian environmental effects would lead to a different fitness function for breeding values). The first problem is to calculate the moments, and hence the cumulants, after selection. This can be done by referring to the definition, $\Psi^*(g) = \Psi(g)W(g)/\bar{W}$, and using the integrals

$$\int_{-\infty}^{\infty} \phi(y)\Phi\left(\frac{y - T}{\alpha}\right)dy = \Phi\left(\frac{-T}{\sqrt{1 + \alpha^2}}\right) \quad (5a)$$

and

$$\int_{-\infty}^{\infty} \phi(y)H_j(y)\Phi\left(\frac{y - T}{\alpha}\right)dy$$
$$(5b)$$
$$= E\left[H_{j-1}\left(\frac{T + \alpha X\sqrt{1 + \alpha^2}}{1 + \alpha^2}\right)\right]\frac{\exp\left(\frac{-T^2}{2(1 + \alpha^2)}\right)}{\sqrt{2\pi(1 + \alpha^2)}},$$

where $E[f(X)]$ is the expectation of the polynomial $f(X)$, with X following a standard normal distribution. Since any polynomial $f(y)$ can be expressed as a sum of Hermite polynomials, Equation 5 can be used to evaluate any integral of the form $\int_{-\infty}^{\infty} \phi(y)f(y)\Phi((y - T)/\alpha)\,dy$.

Reproduction reduces the ith cumulant by a factor $2^{-(i-1)}$ for $i = 2, 3, \ldots$ and adds $V_{G,LE}/2$ to the variance (BULMER 1980, Ch. 9). By setting these cumulants after selection and reproduction equal to their values among zygotes, one obtains numerical solutions for the steady-state rate of advance and for the equilibrium variance and higher-order cumulants under recurrent truncation selection of fixed intensity. The same values can be obtained, at least in principle, by iterating the integral equation (2). At steady state, the cumulants of order two and higher become constant and the mean changes by a constant amount each generation.

Figure 1 shows the rate of advance, additive genetic variance, normalized skew ($\gamma_3 = E[(G - \bar{Z})^3]/V_G^{3/2}$), and kurtosis ($\gamma_4 = E[(G - \bar{Z})^4]/V_G^2 - 3$) of breeding values as a function of the mean fitness (i.e., the proportion selected). To emphasize the relatively small cumulative effects of selection on the higher-order cumulants, the additive variance is adjusted so that the heritability is $h^2 = 0.5$ for both the one-generation and equilibrium calculations. In these figures, the truncation point was calculated using the standard formula from Gaussian theory rather than the more elaborate Cornish-Fisher approximation used by ZENG (1987). The differences in the truncation points obtained for these small values of skew and kurtosis are negligible. Following ZENG (1987), the equilibrium moments were calculated using a fourth-order Gram-Charlier expansion. Table 2 shows that higher cumulants rapidly decline to zero, and that adding extra terms to the expansion makes very little difference. (Note that the Gram-Charlier expansion gives a distribution that becomes negative for large y if it is defined with an odd number of terms, and may do so with an even number. However, this pathology arises only for y well outside the range of interest.)

The dotted curves in Figure 1, c and d, show the distortion produced by one generation of selection from a Gaussian distribution, while the solid curves show the distortion after the population has settled to a steady advance. These curves are close to each other, showing that the population rapidly equilibrates. Surprisingly, the greatest steady-state skew is produced with relatively weak selection, e.g., with $\bar{W} \approx 85\%$ for $h^2 = 0.5$. The
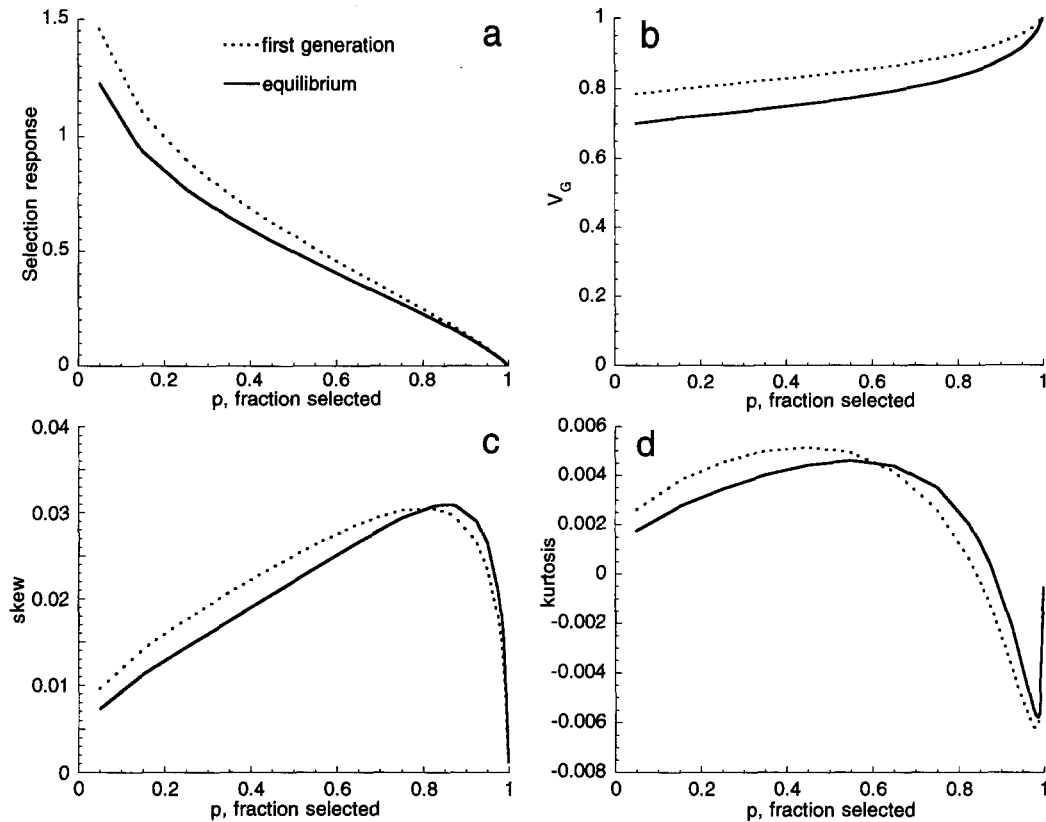
FIGURE 1.—Response to truncation selection under the infinitesimal model. Individual panels display: (a) the change in mean per generation, (b) the equilibrium genetic variance (V_G), (c) the skew $(\gamma_3 = E[(G - \bar{Z})^3]/V_G^{3/2})$, and (d) the kurtosis $(\gamma_4 = E[(G - \bar{Z})^4]/V_G^2 - 3)$, as a function of the proportion selected. All moments are expressed relative to the genic variance, $V_{G,LE}$, that would be reached with no selection. The dotted curves give the values after one generation, starting from a Gaussian with variance $V_{G,LE}$, and the solid curves give the steady-state values. Heritability is $h^2 = 0.5$ for both sets of curves. All results were calculated using a fourth-order Gram-Charlier expansion.

TABLE 2

Effects of increasing the number of terms (imax) in the Gram-Charlier expansion (Equation 3) on the predicted change

imax	R	V_G	Skew	Kurtosis	Pentosis[a]	Hexosis[b]
2	0.2251	0.8275				
3	0.2245	0.8321	0.03022			
4	0.2245	0.8321	0.03037	0.002605		
5	0.2245	0.8321	0.03038	0.002558	−0.002099	
6	0.2245	0.8322	0.03036	0.002552	−0.002022	−0.001203

Changes are shown in the mean per generation (R), the genetic variance $(V_G$, standardized by the genic variance), and the standardized cumulants (defined by $K_i/V_G^{i/2}$ for $i = 3, 4, 5, 6$) under truncation selection with approximately 80% of the population selected [the truncation point is set to $-1.190\sqrt{V_G}$, so that $\bar{W} \cong 0.8$, which produces a relatively large skew (cf. Figure 1)] and $h^2 = 0.5$.

[a] $K_5/V_G^{5/2} = (C_5 - 10C_3C_2)/V_G^{5/2}$, where $C_i = E\{[G - E(G)]^i\}$.

[b] $K_6/V_G^3 = (C_6 - 15C_4C_2 - 10C_3^2 + 30C_2^3)/V_G^3$.

steady-state kurtosis is also largest when selection is weak; it changes from its most positive value at 55% selected to its most negative value when 98.5% are selected. However, the skew and kurtosis never become large, and have no appreciable effect on the response to selection, or on the genetic variance. We will show below that the same small effects can be deduced from a purely genetic, rather than statistical, argument.

Figure 2 shows how the skew and kurtosis change with the heritability, for 20%, 50% and 80% selection; values are calculated using the fourth-order Gram-Charlier expansion. Though deviations from normality increase

with the heritability, they never become large. Even when the trait is completely heritable, the skew never rises above 0.1, and the kurtosis never rises above 0.02.

Using Fourier transforms: In this method, the distribution after selection is found directly, by multiplying the initial distribution by the relative fitness function. Finding the effect of reproduction directly from Equation 2 would be slow, because it requires two integrations for each value of z. However, the Fourier transform of the distribution, $\hat{\psi}(\hat{z})$ (defined in Equation 6b below), can be found rapidly using the fast Fourier transform algorithm. The effect of reproduction on the Fourier
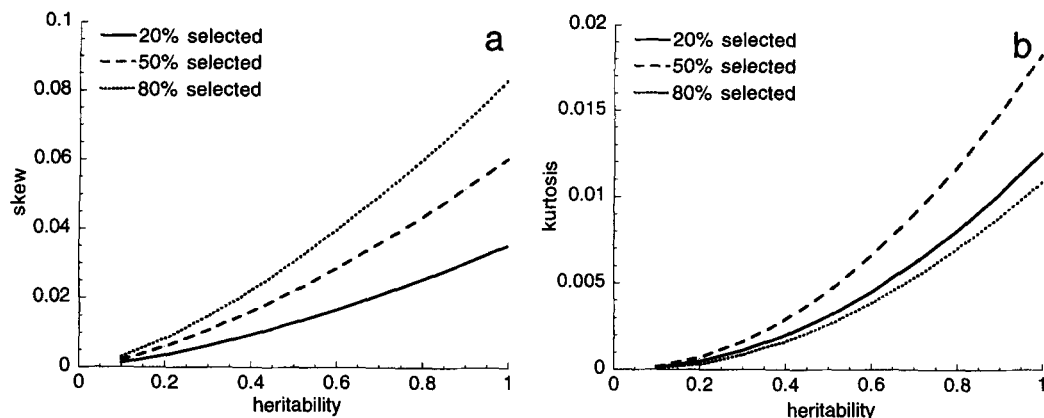
FIGURE 2.—The skew (a) and kurtosis (b) produced under steady truncation selection, as a function of the heritability. Values are calculated as for Figure 1, with proportions selected being 20%, 50% and 80%.

transform (*i.e.*, the convolution Equation 2) is then given by a simple multiplication, so that

$$\Psi^{**}(\tilde{z}) = \Psi^{*}\left(\frac{z}{2}\right)^{2}\sqrt{2\pi}\exp\left(\frac{-\tilde{z}^{2}V_{G,LE}}{4}\right), \quad (6a)$$

where

$$\Psi^{**}(\tilde{z}) = \int_{-\infty}^{\infty}\Psi^{*}(z)\frac{\exp(iz\tilde{z})}{\sqrt{2\pi}}dz. \quad (6b)$$

Taking the inverse Fourier transform completes the calculation,

$$\Psi^{*}(z) = \int_{-\infty}^{\infty}\Psi^{**}(\tilde{z})\frac{\exp(-iz\tilde{z})}{\sqrt{2\pi}}d\tilde{z}. \quad (6c)$$

Numerical results can readily be found by defining the distribution $\Psi(z)$ on an evenly spaced set of points over some finite range, and then applying the fast Fourier transform algorithm (PRESS *et al.* 1989). Unfortunately, this does not give a practical way of calculating the slight deviations from normality caused by truncation selection. The discrete approximation causes slight errors in the tails of the distribution, which though small, cause large errors in the variance and higher moments. For example, suppose that the distribution is initially Gaussian, with mean zero and unit variance. Iterating the algorithm over 1 generation with no selection, and with genic variance $V_{G,LE} = 1$, should not change the distribution. Table 3 shows that although the error in the distribution itself is small, the variance and kurtosis are inaccurate. Since truncation selection causes only slight perturbations away from normality, the Fourier transform method would require an impracticably fine grid to give sufficient accuracy.

The Fourier transform method is more appropriate for cases in which selection generates large deviations from normality; then, the distribution cannot be adequately described in terms of its first few cumulants. To illustrate this, consider an equilibrium under disruptive selection. We follow FELSENSTEIN (1979) and assume that

the fitness of an individual with phenotype Z is the sum of two Gaussians: $W(z) = \exp[-s(z - \theta)^{2}/2] + \exp[-s(z + \theta)^{2}/2]$. With environmental variance V_E, fitness as a function of breeding value G is

$$W^{*}(g) = \frac{\exp[-s^{*}(g - \theta)^{2}/2] + \exp[-s^{*}(g + \theta)^{2}/2]}{\sqrt{1 + sV_E}}, \quad (7)$$

where $s^{*} = s/(1 + sV_E)$; this is just a smoothed version of the individual fitness.

For simplicity, we consider only the symmetric equilibria at which the population mean is zero. FELSENSTEIN (1979) showed, using a multivariate normal model with constant fitnesses, that these equilibria are unstable when disruptive selection is strong enough that $W^{*}(g)$ is bimodal (*cf.* BULMER 1980, Ch. 10). However, if the two peaks correspond to two different limiting resources, frequency-dependent selection can act to keep the mean at zero. [FELSENSTEIN (1979) showed that the symmetric equilibria were only quadratically unstable; thus, weak frequency-dependence may suffice.] Although these equilibria are unstable under our constant fitness model, they can be approximated by iterating the recursions beginning with the population mean at zero.

When selection is weak, $W^{*}(g)$ has a single peak (corresponding to stabilizing selection); one expects an approximately Gaussian distribution, with variance slightly lower than the genic variance, $V_{G,LE}$. As selection becomes stronger, $W^{*}(g)$ develops two peaks near $\pm\theta$, and the variance should be inflated above $V_{G,LE}$. When selection is very strong, only individuals near $\pm\theta$ will survive. After these mate at random, $\frac{1}{4}$ of the population will consist of offspring distributed with variance $V_{G,LE}/2$ around $-\theta$, $\frac{1}{4}$ will be distributed with variance $V_{G,LE}/2$ around $+\theta$, and $\frac{1}{2}$ will be distributed with variance $V_{G,LE}/2$ around zero. Thus, strong disruptive selection can produce a distribution of breeding values that is approximately a mixture of three Gaussians. These expectations are confirmed by the results in Figure 3. The lower panels show a transition from an approximately

TABLE 3

Values produced by the Fourier transform method after one generation without selection, starting from a Gaussian distribution with genetic variance and genic variance equal to 1, using grids of different mesh and different ranges of integration

h^a	Variance			Kurtosis			$\Delta\Psi_{max}{}^c$
	$(-8,8)^b$	$(-6,6)$	$(-4,4)$	$(-8,8)$	$(-6,6)$	$(-4,4)$	
32	1.0266	0.9872	1.0027	1.0465	−0.8538	−0.3866	0.022555
64	0.9805	0.9948	0.9971	−0.7535	−0.1043	−0.0361	0.001053
128	0.9829	0.9953	0.9968	−0.6224	−0.0689	−0.0160	0.000633
256	0.9842	0.9955	0.9968	−0.5684	−0.0608	−0.0148	0.000732

^{a} The number of points spaced over the range $(-8, 8)$ used for the fast Fourier transform.
^{b} The range over which the moments were calculated by numerical integration.
^{c} The maximum deviation of the numerically determined probability density function from the expected Gaussian.
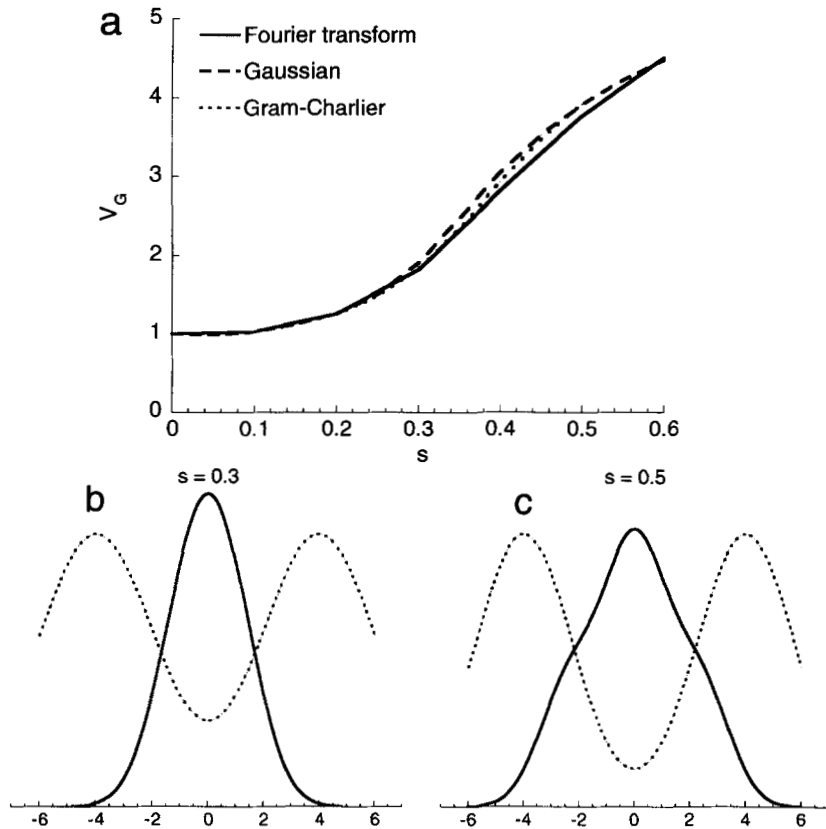

FIGURE 3.—Consequences of disruptive selection under the infinitesimal model. Graph a shows the genetic variance maintained. The solid curve gives values calculated using Fourier transforms. The algorithm was iterated for 10 generations, starting from a Gaussian with variance $V_G = V_E$. The variance changed little after the first 5 generations, so the generation 10 values are assumed to be near equilibrium. The dashed curve assumes a Gaussian distribution and the dotted curve assumes a fourth-order Gram-Charlier expansion. Graphs b and c show the distribution for $s = 0.3$ and $s = 0.5$, calculated using Fourier transforms (solid curves), together with the graph of fitness as a function of breeding value (dashed curves). The fitness of individuals with phenotype x is $W(x) = \exp(-s(x - \theta)^2/2) + \exp(-s(x + \theta)^2/2)$, where $\theta = 4$; the environmental variance is $V_E = 1$, and the genic variance is $V_{G,LE} = 1$.

Gaussian to a distinctly non-Gaussian distribution; for $\theta = 4$ and $V_{G,LE} = V_E = 1$, the transition occurs around $s = 0.4$. However, even though the distribution is far from Gaussian, the variance predicted assuming a Gaussian distribution [cf. BULMER (1980, Ch. 9)] is close to that calculated using the Fourier transform (dashed vs. solid line in Figure 3a). Surprisingly, the Gram-Charlier method does not much improve the fit: adding a fourth-order term makes little difference (dotted line in Figure 3a). Higher order Gram-Charlier calculations are intractable. Thus, even when selection maintains a substantially non-Gaussian distribution, the Gaussian approximation for the variance is quite good, and including further cumulants does not much improve it. An alternative, population-genetic analysis of disruptive selection is presented below.

Our analysis of disruptive selection is fundamentally different from FELSENSTEIN's. We are concerned with the short-term changes in the distribution caused by linkage disequilibria, and take the genic variance (contributed by heterozygosity at individual loci) to be fixed. On a longer time scale, we expect that if selection is strong enough, the genic variance will increase until it approaches the equilibrium identified by FELSENSTEIN (1979): $V_G = \theta^2 - (1 + sV_E)/s$. This is the point where selection does not alter the variance (in terms of the selection gradient notation introduced below this corresponds to $\mathcal{L}_2 = 0$). The stability of the equilibrium depends on the details of the genetics, and on the degree of frequency-dependent selection; further analysis is needed to confirm our intuition that equilibrium will be reached when $\mathcal{L}_2 = 0$.

GENETIC ANALYSIS

A full description of the dynamics of the allele frequencies and linkage disequilibria that determine the distribution of the trait is more complicated than these statistical analyses. Our approach is based on WRIGHT's (1935) formula, in which the change in gamete frequencies caused by selection is exactly proportional to the gradient of log mean fitness with respect to those gamete frequencies. BARTON and TURELLI (1987) showed how Wright's equation could be rewritten in terms of any other description of the population–for example, the mean and central moments of the genotypic distribution. Since this rotation of coordinates is in general nonlinear, it is accurate only to first order in selection. TURELLI and BARTON (1990) used this method to approximate the joint effects of weak selection with recombination, and BARTON and TURELLI (1991, Equation 22) showed how the exact equations under strong selection can be derived from the first-order expressions valid for weak selection. The latter paper described selection by representing the individual relative fitness ($W/\bar{W}$) as a polynomial. Here, however, we describe selection in terms of gradients in mean fitness.

Describing the population in terms of cumulants: Our treatment below differs from our previous analyses mainly in that we describe the population in terms of multilocus *cumulants, K,* rather than the *mean m* and the *central moments C.* This approach was applied by BÜRGER (1991), in his analysis of weak selection on a continuum-of-alleles; he also introduced the use of a generating function to calculate the selection equations. We extend BÜRGER's results first by allowing for strong selection of arbitrary form, and second by allowing for linkage disequilibria among an arbitrary number of loci.

The cumulants are a set of parameters that describe the shape of a probability distribution. The first three equal the mean, the variance, and the third central moment, while the higher cumulants are polynomial functions of the moments. All cumulants have the convenient additivity properties of the mean and the variance (see Equation 12 below). This is particularly useful in analyzing additive polygenic traits: for example, the gradients of log mean fitness with respect to the cumulants of effects at individual loci are the same as the gradients with respect to the cumulants of the overall trait distribution. Expressions for the response of the cumulants to selection also simplify the equations for arbitrary multilocus selection. A further advantage is that for a normal distribution, the third and higher-order cumulants are all zero. We therefore expect that if the trait is normally distributed, its response to selection can be explained solely in terms of the selection gradients with respect to the mean and variance.

Cumulants bring two disadvantages. First, while the first-order expressions for the response of the non-

central moments to selection are exact and apply for arbitrary selection strength, those for the cumulants are not. Second, the effects of recombination are most naturally described in terms of the moments (TURELLI and BARTON 1990). However, it is easy to move between these alternative representations. We therefore use cumulants to describe selection, but follow changes of the moments under selection and recombination, then convert these to changes in the cumulants.

For a single random variable X, the relations between the cumulants, K_i, and the first few moments, $m = E[X]$ and $C_i = E[(X - m)^i]$, are

$$K_1 = m, \tag{8a}$$

$$K_2 = C_2, \tag{8b}$$

$$K_3 = C_3, \tag{8c}$$

and
$$K_4 = C_4 - 3C_2^2. \tag{8d}$$

The general definition is best given in terms of the moment generating function. Let $C_i^* = E(X^i)$ denote the ith non-central moment. For probability density $f(x)$, the moment generating function is

$$\hat{f}(\hat{x}) = \int_{-\infty}^{\infty} \exp(x\hat{x})f(x)\, dx, \tag{9a}$$

and
$$C_i^* = \frac{\partial^i \hat{f}(0)}{\partial \hat{x}^i}. \tag{9b}$$

(The moment generating function is equivalent to the Fourier transform: $\hat{f}(\hat{x}) = \tilde{f}(-i\hat{x})\sqrt{2\pi}$.) The cumulants are given by the derivatives of the cumulant generating function, $K(\hat{x})$, which is just the natural log of the moment generating function, *i.e.*

$$K(\hat{x}) = \ln[\hat{f}(\hat{x})] \tag{10a}$$

and
$$K_i = \frac{\partial^i K(0)}{\partial \hat{x}^i}. \tag{10b}$$

The multivariate cumulants are defined in the same way (STUART and ORR 1987, Ch. 3). Now, **X** is a random vector with probability density $f(\mathbf{x})$. The moment generating function is

$$\hat{f}(\hat{\mathbf{x}}) = \int_{-\infty}^{\infty} \exp(\mathbf{x}\cdot\hat{\mathbf{x}})f(\mathbf{x})\, d\mathbf{x}$$
$$\tag{11a}$$
$$\text{with} \quad \mathbf{x}\cdot\hat{\mathbf{x}} = \sum_i x_i\, \hat{x}_i,$$

and
$$C_U^* = \frac{\partial^{|U|}\hat{f}(0)}{\partial \hat{\mathbf{x}}_U} \tag{11b}$$

$$\text{where} \quad C_U^* = E(\mathbf{X}_U) \quad \text{and} \quad \mathbf{X}_U = \prod_{i\in U} X_i.$$

Here, $\partial^{|U|}\hat{f}/\partial\hat{\mathbf{x}}_U$ denotes the partial differential with

respect to the variables $\hat{x}_i$ in the set U (see Table 1); and C_U^* denotes the expectation (E) with respect to $f(\mathbf{x})$ across a set of indices. For example, $C_{\{iijk\}}^* \equiv C_{iijk}^* = E(X_i^2 X_j X_k) = \partial^4 \hat{f}(0)/\partial x_i^2 \partial x_j \partial x_k$. As before, the multivariate cumulants are given by the derivatives of the cumulant generating function, $K(\hat{\mathbf{x}})$, with

$$K(\hat{\mathbf{x}}) = \ln[\hat{f}(\hat{\mathbf{x}})] \tag{12a}$$

and

$$K_U = \frac{\partial^{|U|} K(0)}{\partial \hat{\mathbf{x}}_U}. \tag{12b}$$

As noted above, cumulants have a convenient additivity property. Suppose that $Z = \sum_i X_i$, where $(X_1, X_2, X_3, \ldots)$ has multivariate cumulants denoted κ_{ijk}, etc. Then K_j, the jth cumulant of the sum Z, is simply related to the multivariate cumulants by

$$K_j = \sum_{U:|U|=j} \kappa_U, \tag{12c}$$

where $|U|$ denotes the number of elements in the set U. For instance, if $Z = X_1 + X_2$, $K_2 = \kappa_{11} + \kappa_{22} + 2\kappa_{12}$ (since $\kappa_{12} = \kappa_{21}$). Equation 12c follows from the fact that the univariate moment generating function $\hat{f}_Z(\hat{x})$ for Z is just $\hat{f}_{\mathbf{X}}(\hat{x}, \hat{x}, \ldots, \hat{x})$, where $\hat{f}_{\mathbf{X}}$ denotes the multivariate moment generating function for $\mathbf{X}$. However, we stress that our general treatment does not assume additivity of genetic effects.

We will be concerned with the distribution of n-locus diploid genotypes, $f(\mathbf{x}, \mathbf{x}^*)$. Here, $\mathbf{x}$ is a vector labeling the alleles derived from the mother, and $\mathbf{x}^*$ the corresponding vector for alleles from the father. With random mating, $\mathbf{x}$ and $\mathbf{x}^*$ are independent in zygotes; but, selection usually generates deviations from multilocus Hardy-Weinberg proportions, necessitating analysis of the joint distribution $f(\mathbf{x}, \mathbf{x}^*)$. However, for compactness, we will set out the methods in terms of a single vector $\mathbf{x}$, which we take to include the states of the full diploid genotype. The expansion to diploid notation is trivial, as indicated in the final equations below (see 19).

The multilocus cumulants defined by Equations 11 and 12 are a new measure of deviation from linkage equilibrium. When more than three loci are involved, they are distinct from those based on the cross-locus moments defined by SLATKIN (1972, Equation 6) and used in our previous papers. They are also distinct from BENNETT's (1954) principal components, which were defined so as to decline geometrically in the absence of selection. Though cumulants are more convenient for analyzing selection, they are less so for recombination. To see this, consider κ_{ijkl}^* the four-locus cumulant after recombination alone. (In terms of central moments, denoted C_N, (12) implies that $\kappa_{ijkl} = C_{ijkl} - C_{ij}C_{kl} - C_{ik}C_{jl} - C_{il}C_{jk}$, where C_{ij} denotes the covariance of allelic effects between loci i and j.) For simplicity, we assume that selection has generated no associations between paternally and maternally derived loci, so that $K_{U,V} = 0$ if U and V are non-empty sets. As in BARTON and TURELLI (1991), we define $r_{S,T}$ for nonoverlapping sets of indices S and T as the probability of recombination events

that bring together paternally derived alleles from the loci in S with maternally derived alleles from the loci in T; similarly, $r_{N,\phi}$ is the frequency of gametes that are "non-recombinant" with respect to all loci in N. By converting to central moments, it is easy to show that

$$\begin{aligned} \kappa_{ijkl}^* = {} & r_{ijkl,\varnothing}\kappa_{ijkl} + (r_{ijkl,\varnothing} + r_{ij,\,kl} - r_{ij,\varnothing}r_{kl,\varnothing})\kappa_{ij}\kappa_{kl} \\ & + (r_{ijkl,\varnothing} + r_{ik,\,jl} - r_{ik,\varnothing}r_{jl,\varnothing})\kappa_{ik}\kappa_{jl} \\ & + (r_{ijkl,\varnothing} + r_{il,\,jk} - r_{il,\varnothing}r_{jk,\varnothing})\kappa_{il}\kappa_{jk}. \end{aligned} \tag{13}$$

Here, $r_{ij,\varnothing} \equiv 1 - r_{ij}$, where r_{ij} is the recombination rate between loci i and j. The factor $(r_{ijkl,\varnothing} + r_{ij,\,kl} - r_{ij,\varnothing}r_{kl,\varnothing})$ is a measure of association between the pair of loci $\{ij\}$ and the pair $\{kl\}$; it is zero if the event that $\{ij\}$ stay together is independent of the event that $\{kl\}$ stay together. That will be true if the pairs do not overlap on the same chromosome (e.g., if they are in the order (i, j, k, l) but not (i, k, j, l)). Only if all the loci are unlinked will the cumulant measures of disequilibrium be identical with those of BENNETT (1954).

Response to strong selection: The exact change in genotype frequencies caused by selection is

$$\Delta_s f(\mathbf{x}) = \frac{f(\mathbf{x})(W(\mathbf{x}) - \bar{W})}{\bar{W}} = \int G[\mathbf{x}; \mathbf{y}] \frac{\partial \ln(\bar{W})}{\partial f(\mathbf{y})} d\mathbf{y}, \tag{14a}$$

with

$$G[\mathbf{x}; \mathbf{y}] = f(\mathbf{x})[\delta(\mathbf{x} - \mathbf{y}) - f(\mathbf{y})], \tag{14b}$$

where $\delta(\mathbf{x} - \mathbf{y})$ is the Dirac delta function and $\bar{W} = \int f(\mathbf{x})W(\mathbf{x}) d\mathbf{x}$ (see Equations 2.2 and 2.3 of TURELLI and BARTON 1990). With discrete alleles, the integral in (14a) would be replaced by a sum and the Dirac delta in (14b) would be replaced by the Kronecker delta. The crucial step is to describe the population in terms of the moment generating function $\hat{f}(\hat{\mathbf{x}})$ rather than the genotype frequencies. Using the transformation methods described by Equations 2.4 and 2.5 of TURELLI and BARTON (1990), we find that, because $\hat{f}(\hat{\mathbf{x}})$ is a linear function of $f(\mathbf{x})$ (see Equation 9),

$$\Delta_s \hat{f}(\hat{\mathbf{x}}) = \int \hat{G}[\hat{\mathbf{x}}; \hat{\mathbf{y}}] \frac{\partial \ln(\bar{W})}{\partial \hat{f}(\hat{\mathbf{y}})} d\hat{\mathbf{y}}, \tag{15a}$$

with

$\hat{G}[\hat{\mathbf{x}}; \hat{\mathbf{y}}]$

$$= \int \int \frac{\partial \hat{f}(\hat{\mathbf{x}})}{\partial f(\hat{\mathbf{x}})} f(\mathbf{x})[\delta(\mathbf{x} - \mathbf{y}) - f(\mathbf{y})] \frac{\partial \hat{f}(\hat{\mathbf{y}})}{\partial f(\hat{\mathbf{y}})} d\mathbf{x} \, d\mathbf{y}$$

$$= \int \int \exp(\mathbf{x} \cdot \hat{\mathbf{x}}) f(\mathbf{x})[\delta(\mathbf{x} - \mathbf{y}) - f(\mathbf{y})] \exp(\mathbf{y} \cdot \hat{\mathbf{y}}) \, d\mathbf{x} \, d\mathbf{y}$$

$$= \hat{f}(\hat{\mathbf{x}} + \hat{\mathbf{y}}) - \hat{f}(\hat{\mathbf{x}})\hat{f}(\hat{\mathbf{y}}). \tag{15b}$$

Because the relation between the moment and cumulant generating functions is not linear, the expression corresponding to (15) for the change in the cumulant generating function is accurate only to first order in the selection gradients (denoted by $\mathcal{L}$ below).

Because recombination produces a linear mixture of probability densities, its effects are simplest to describe for statistics that depend linearly on the density. Thus, we will describe the effects of selection and recombination on moments rather than cumulants. However, because of the additivity property (12), it is simpler to work with selection gradients defined in terms of the cumulants, rather than the moments (see Appendix 1 of TURELLI and BARTON 1990). To achieve this, we replace $\partial \ln(\bar{W})/\partial \hat{f}(\hat{y})$ in (15a) by $[\partial \ln(\bar{W})/\partial K(\hat{y})][\partial K(\hat{y})/\partial \hat{f}(\hat{y})]$. The resulting expression for the moments is exact, but is somewhat less elegant than the analogous (first order) equation for cumulants (B1), since it involves an asymmetric matrix G^*. In terms of generating functions, we have

$$\Delta_s \hat{f}(\hat{x}) = \int \hat{G}^*[\hat{x}; \hat{y}] \frac{\partial \ln(\bar{W})}{\partial K(\hat{y})} \, d\hat{y}, \qquad (16a)$$

with

$$\hat{G}^*[\hat{x}; \hat{y}] = [\hat{f}(\hat{x} + \hat{y}) - \hat{f}(\hat{x})\hat{f}(\hat{y})] \frac{\partial K(\hat{y})}{\partial \hat{f}(\hat{y})}$$

$$= [\hat{f}(\hat{x} + \hat{y}) - \hat{f}(\hat{x})\hat{f}(\hat{y})] \frac{\partial \ln[\hat{f}(\hat{y})]}{\partial \hat{f}(\hat{y})}$$

$$= \frac{\hat{f}(\hat{x} + \hat{y})}{\hat{f}(\hat{y})} - \hat{f}(\hat{x}) \qquad (16b)$$

$$= \exp[K(\hat{x} + \hat{y}) - K(\hat{y})] - \exp[K(\hat{x})].$$

As explained below, the matrix G^* that gives the response of particular non-central moments C_U^* to selection gradients on particular multilocus cumulants, denoted κ_V, can be found by differentiating the generating function $\hat{G}^*[\hat{x}; \hat{y}]$ to produce

$$\Delta_s C_U^* = \sum_V G^*[U; V] \mathcal{L}_V, \qquad (17a)$$

where

$$\mathcal{L}_V = \frac{\partial \ln(\bar{W})}{\partial \kappa_V} \quad \text{and} \quad G^*[U; V] = \frac{\partial^{|U+V|} \hat{G}^*[0; 0]}{\partial \hat{x}_U \partial \hat{y}_V}. \qquad (17b)$$

Here, $\Delta_s C_U^*$ is obtained from $\Delta_s \hat{f}(\hat{x})$ by simply differentiating with respect to $\hat{x}_U$ (*i.e.*, with respect to x_i for all loci $i \in U$) and evaluating the result at $\hat{x} = \mathbf{0}$. It is less obvious that the effect of selection on particular cumulants, $\mathcal{L}_V$, is obtained by differentiating with respect to $\hat{y}_V$. This can be verified by working backward from (17a):

$$\Delta_s C_U^* = \sum_V \frac{\partial^{|U+V|} \hat{G}^*[0; 0]}{\partial \hat{x}_U \partial \hat{y}_V} \frac{\partial \ln(\bar{W})}{\partial \kappa_V}$$

$$\qquad (18a)$$

$$= \int \sum_V \frac{\partial^{|U+V|} \hat{G}^*[0; 0]}{\partial \hat{x}_U \partial \hat{y}_V} \frac{\partial \hat{K}(\hat{w})}{\partial \kappa_V} \frac{\partial \ln(\bar{W})}{\partial \hat{K}(\hat{w})} \, d\hat{w}.$$

Because $\hat{K}(\hat{w}) = \sum_V (\hat{w}_V / |V|!) \kappa_V$ (see Equations 11 and 12), (18a) becomes

$$\Delta_s C_U^* = \int \sum_V \frac{\partial^{|U+V|} \hat{G}^*[0; 0]}{\partial \hat{x}_U \partial \hat{y}_V} \frac{\hat{w}_V}{|V|!} \frac{\partial \ln(\bar{W})}{\partial \hat{K}(\hat{w})} \, d\hat{w}. \qquad (18b)$$

The sum over all V gives a Taylor series in $\hat{w}$; hence, (18b) can be written as

$$\Delta_s C_U^* = \int \frac{\partial^{|U|} \hat{G}^*[0; \hat{w}]}{\partial \hat{x}_U} \frac{\partial \ln(\bar{W})}{\partial \hat{K}(\hat{w})} \, d\hat{w}. \qquad (18c)$$

From (11b), we see that (18c) reproduces (16a), and so verifies Equation 17.

The expressions for G^* become complicated for the higher cumulants; for example, $G^*[\{i, j, k, l\}; \{m, n, o, p\}]$ involves 339 terms. However, Equations 16 and 17 give a compact algorithm that can readily be automated, using a symbolic language such as *Mathematica* (WOLFRAM 1991). (A description of the routines used to perform these calculations as well as *Mathematica* notebooks for the Macintosh that implement them are available upon request.) The expression for changes in the cross-locus moments is essentially the same. It is obtained by replacing $\hat{x}$ by $(\hat{x}, \hat{x}^*)$ and $\hat{y}$ by $(\hat{y}, \hat{y}^*)$, which produces

$$\Delta_s C_{U,V}^* = \sum_{S,T} G^*[U, V; S, T] \mathcal{L}_{S,T}$$

$$\qquad (19)$$

$$\text{where} \quad \mathcal{L}_{S,T} = \frac{\partial \ln(\bar{W})}{\partial \kappa_{S,T}},$$

$$G^*[U, V; S, T] = \frac{\partial^{|U+V+S+T|} \hat{G}^*[0, 0; 0, 0]}{\partial \hat{x}_U \partial \hat{x}_V^* \partial \hat{y}_S \partial \hat{y}_T^*}, \qquad (19b)$$

and

$$\hat{G}^*[\hat{x}, \hat{x}^*; \hat{y}, \hat{y}^*] \qquad (19c)$$

$$= \exp[K(\hat{x} + \hat{y}, \hat{x}^* + \hat{y}^*) - K(\hat{y}, \hat{y}^*)] - \exp[K(\hat{x}, \hat{x}^*)].$$

Equation 19 gives the response of the moments to selection on the cumulants. BÜRGER (1991, Equations 4.4b and 4.6) gives closely related expressions for the change in cumulants, which are also derived from a generating function. His results differ in that they are expressed in terms of the coefficients of a Taylor series expansion of the fitness as a function of breeding value (his Equation 3.5), rather than selection gradients. His results also neglect linkage disequilibrium, and so are the sum of a set of single-locus analyses. Our equations are essentially an extension of his analysis to arbitrary multilocus systems and to strong selection.

Allowing for recombination: The exact effects of strong selection on the non-central moments are given by Equation 19. This equation makes no assumptions about the mapping of genotypes onto phenotypes. It requires only that the distribution of genotypes can be specified in terms of moments. As discussed in BARTON and TURELLI (1991), this is straightforward for diallelic loci or models with additive allelic effects. Recombination is then dealt with using Equations 14 and 15 of BARTON and TURELLI (1991). Finally, the non-central moments after selection and recombination are used to calculate the new cumulants. This rather complex algorithm is justified because cumulants are a natural way to

describe deviations from normality, and they simplify the characterization of the distributions of breeding values and phenotypes under the additive model.

Calculating the selection gradients under an additive model: We will now restrict attention to a simple additive model for polygenic inheritance. Letting Z denote the phenotype of a randomly chosen individual, we assume

$$Z = \sum_{i=1}^{n} (X_i + X_i^*) + \mathscr{E} \equiv G + \mathscr{E}, \qquad (20)$$

where $X_i(X_i^*)$ denotes the additive contribution of the maternally (paternally) inherited allele at autosomal locus i, G denotes the total genetic contribution, and $\mathscr{E}$ denotes a random, Gaussian-distributed environmental effect with mean 0 and variance V_E, which is independent of G. Under this model, K_j, the jth cumulant of the distribution of phenotypes, is the sum of the jth order cumulants across all sets of j loci (plus V_E for the second cumulant) (see Equation 12). Hence, the selection gradient with respect to any cumulant involving j loci is the same as the selection gradient with respect to the jth cumulant of the genotypic (or phenotypic) distribution. In particular, the terms $\mathscr{L}_{S,T}$ in Equation 19a reduce to $\mathscr{L}_j = \partial \ln(\bar{W})/\partial K_j$ with $j = |S + T|$, the total number of indices in the sets S and T. (This reduction occurs with any distribution of environmental deviations, $\mathscr{E}$, as long as $\mathscr{E}$ is independent of G.) We now describe how these genotypic gradients can be calculated.

We first find how the distribution of phenotypes, $f(x)$, depends on K_j. The Fourier transform of a distribution, $\tilde{f}$ (Equation 6b), is equivalent to its moment generating function, $\hat{f}$ (Equation 8a), which in turn can be written as a Taylor series in the cumulants (Equation 9), *i.e.*,

$$\tilde{f}(\tilde{x}) = \frac{1}{\sqrt{2\pi}} \hat{f}(i\tilde{x}) = \frac{1}{\sqrt{2\pi}} \exp[K(i\tilde{x})]$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left[\sum_{j=0}^{\infty} K_j \frac{(i\tilde{x})^j}{j!} \right]. \qquad (21)$$

Hence, $\qquad \dfrac{\partial \tilde{f}(\tilde{x})}{\partial K_j} = \dfrac{(i\tilde{x})^j}{j!} \tilde{f}(\tilde{x}). \qquad (22)$

Taking the inverse Fourier transform (Equation 6c) of Equation 22,

$$\frac{\partial f(x)}{\partial K_j} = \int_{-\infty}^{\infty} \frac{\partial \tilde{f}(\tilde{x})}{\partial K_j} \frac{\exp(-ix\tilde{x})}{\sqrt{2\pi}} d\tilde{x}$$

$$= \int_{-\infty}^{\infty} \frac{(i\tilde{x})^j}{j!} \tilde{f}(\tilde{x}) \frac{\exp(-ix\tilde{x})}{\sqrt{2\pi}} d\tilde{x} \qquad (23)$$

$$= \frac{(-1)^j}{j!} \frac{d^j f(x)}{dx^j}.$$

This simple relation can now be used to find the selection gradient,

$$\mathscr{L}_j \equiv \frac{\partial \ln(\bar{W})}{\partial K_j} = \int_{-\infty}^{\infty} \frac{W(x)}{\bar{W}} \frac{\partial f(x)}{\partial K_j} dx$$

$$= \int_{-\infty}^{\infty} \frac{W(x)}{\bar{W}} \frac{(-1)^j}{j!} \frac{d^j f(x)}{dx^j} dx. \qquad (24)$$

This can be interpreted in two ways. First, consider how the mean fitness changes when the distribution is translated by τ, so that $f(x)$ becomes $f(x - \tau)$, and $\partial f/\partial x = -\partial f/\partial \tau$. Setting

$$\bar{W}(\tau) = \int_{-\infty}^{\infty} W(x) f(x - \tau) dx, \qquad (25a)$$

we have

$$\mathscr{L}_j \equiv \frac{\partial \ln(\bar{W})}{\partial K_j} = \frac{1}{j!} \int_{-\infty}^{\infty} \frac{W(x)}{\bar{W}} \left(\frac{\partial^j f(x - \tau)}{\partial \tau^j} \bigg|_{\tau = 0} \right) dx$$

$$= \frac{1}{j!} \frac{\partial^j \ln[\bar{W}(0)]}{\partial \tau^j}. \qquad (25b)$$

Thus, the selection gradient with respect to the jth cumulant is proportional to the jth differential of the log mean fitness with respect to translation of the distribution. Another relation can be found by integrating Equation 24 by parts j times:

$$\mathscr{L}_j \equiv \frac{\partial \ln(\bar{W})}{\partial K_j} = \frac{1}{j!} \int_{-\infty}^{\infty} \frac{f(x)}{\bar{W}} \frac{d^j W(x)}{dx^j} dx. \qquad (26)$$

Thus, the selection gradient $\mathscr{L}_j$ is also proportional to the expectation of the jth differential of the relative fitness. Note that (25b) and (26) are generalizations of the formulas provided by LANDE (1976) and LANDE and ARNOLD (1983) for the "directional selection gradient," $\beta = \partial \ln(\bar{W})/\partial \bar{Z}$, calculated under the assumption that the distribution of phenotypes remains precisely Gaussian (so that τ is just $\bar{Z}$). In this case, $\mathscr{L}_1 = \beta$. Similarly, if $f(x)$ is Gaussian, Equation 26 implies that $\mathscr{L}_2 = \gamma/2$, where γ is the univariate version of the "stabilizing selection gradient" defined by LANDE and ARNOLD (1983, Equation 14b; *cf.* MITCHELL-OLDS and SHAW 1987).

Approximating the selection gradients: Calculations for specific models are simplified if one can characterize the distribution of breeding values (at least approximately) in terms of a finite set of cumulants. This can always be done for a finite number of loci and alleles. For example, with biallelic loci, the distribution of genotypes is completely characterized by cross-locus cumulants involving distinct loci (κ_{ijk}, etc.) and the allele frequencies (Equation 5 in BARTON and TURELLI 1991). If the distribution of breeding values is approximated by some class of density functions, such as the Gram-Charlier, the selection gradients can be approximated using Equation 26. For densities in which the cumulants appear as

parameters, one can also approximate the $\mathcal{L}_i$ "parametrically" by first differentiating the density with respect to the cumulant then integrating. This will not generally give the same result as Equation 26, because as the lower cumulants vary, the higher cumulants change with them, in a way that depends on the class of distributions considered. For small departures from normality, we will show that the $\mathcal{L}_i$ can be well approximated from Equation 26 assuming that $f(x)$ is Gaussian.

Another approach is to derive the selection gradients from a polynomial approximation to the fitness function. This is the method used by BARTON and TURELLI (1991); there was a slight error in the relation between the polynomial coefficients and the gradients given in Appendix A of that article, which is corrected in APPENDIX A below. This polynomial approximation is also related to BÜRGER's (1991) Taylor series expansion of fitness. If the genotypic (or phenotypic) fitness function is a polynomial, we can directly express the mean fitness as a function of the cumulants of the distribution, then differentiate to obtain the $\mathcal{L}_i$. For instance, if the genotypic fitness function is the quartic,

$$W(x) = 1 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4, \qquad (27)$$

then

$$\bar{W} = W(\bar{Z}) + V_G(b_2 + 3b_3\bar{Z} + 6b_4\bar{Z}^2) \\ + K_3(b_3 + 4b_4\bar{Z}) + b_4(K_4 + 3V_G^2), \qquad (28)$$

$$\bar{W}\mathcal{L}_1 = \frac{dW(\bar{Z})}{d\bar{Z}} + V_G(3b_3 + 12b_4\bar{Z}) + 4b_4 K_3, \qquad (29a)$$

$$\bar{W}\mathcal{L}_2 = b_2 + 3b_3\bar{Z} + 6b_4(\bar{Z}^2 + V_G), \qquad (29b)$$

$$\bar{W}\mathcal{L}_3 = b_3 + 4b_4\bar{Z},$$

and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad (29c)$

$$\bar{W}\mathcal{L}_4 = b_4.$$

Note that if we were working with central moments rather than cumulants, the partial derivative of $\ln(\bar{W})$ with respect to the second moment would differ from (29b) because the coefficient of b_4 in (28) would be simply $C_4 = E[(Z - \bar{Z})^4]$, and hence would not contribute.

If $W(x)$ is not a polynomial, it will generally not be possible to calculate the $\mathcal{L}_i$ exactly without assuming a distribution of breeding values. However, many selection regimes, including truncation selection, can be adequately approximated by a quartic over a range of genotypes that includes most of the population; so that $\mathcal{L}_i$ for $i = 1-4$ often suffice for approximating the dynamics. According to the Gaussian infinitesimal model, $\mathcal{L}_1$ and $\mathcal{L}_2$ suffice. We will use a fourth-order description to approximate the multilocus effects of truncation selection. We will show that under truncation selection, and presumably most forms of directional and stabilizing selection, the cumulative effects of third- and higher-order disequilibria are negligible. This is not true, however, for

strong disruptive selection.

Several criteria are possible for approximating an arbitrary fitness scheme by a polynomial, such as minimizing the "distance" between the exact and approximate fitness functions. To predict selection response, however, it seems more natural to approximate the distortion of the distribution of breeding values or phenotypes. This is most easily accomplished by considering how selection alters the non-central moments. Let C_i^* denote the ith non-central moment before selection, and let C_i^{**} denote its value after selection. If the fitness function is

$$W(x) = \sum_{i=0}^{n} b_i x^i, \qquad \text{with} \quad b_0 = 1, \qquad (30)$$

then

$$\bar{W} = \sum_{i=0}^{n} b_i C_i^* \qquad \text{and} \qquad \bar{W}C_i^{**} = \sum_{j=0}^{n} b_j C_{i+j}^*. \qquad (31)$$

Hence if we know how a selection regime changes the non-central moments, we can use (31) to find a polynomial of order n that produces the same changes in the first n moments (and cumulants). Let $\Delta_s C_i^* = C_i^{**} - C_i^*$, $\mathbf{b} = (b_1, b_2, \ldots, b_n)^T$, and $\Delta_s \mathbf{C}^* = (\Delta_s C_1^*, \Delta_s C_2^*, \ldots, \Delta_s C_n^*)^T$, and define the $n \times n$ matrix $A = (a_{ij})$ by $a_{ij} = C_{i+j}^* - C_i^* C_j^{**}$ (the superscript T denotes transpose). Then the coefficients of the approximating polynomial are

$$\mathbf{b} = A^{-1}\Delta_s \mathbf{C}^*. \qquad (32)$$

Under the Gaussian infinitesimal model, changes in the first two moments completely characterize the selection regime, so a quadratic approximation suffices, as noted by LANDE and ARNOLD (1983). For non-Gaussian distributions, changes in the higher moments are also relevant. These may be estimated empirically. In general, however, one must know the initial distribution of breeding values to predict the effects of a given $W(x)$. Because truncation selection produces only small departures from normality, it suffices (as shown numerically below) to consider how selection alters the distribution assuming that it is initially Gaussian. If only individuals with phenotypes $Z > t$ survive, the genotypic fitness function is

$$W(x) = \Phi\left(\frac{x - t}{\sigma_E}\right), \qquad (33)$$

where $\Phi(x)$ is the cumulative normal density and σ_E is the environmental standard deviation. Assuming the current mean of the population is $\bar{Z}$ and its phenotypic standard deviation is σ_P, let p denote the fraction of the population that survives truncation selection, let $t^* = (t - \bar{Z})/\sigma_P$ denote the normalized truncation point, and let $I(p) = \exp[-(t^*)^2/2]/(p\sqrt{2\pi})$ denote the "selection intensity" (FALCONER
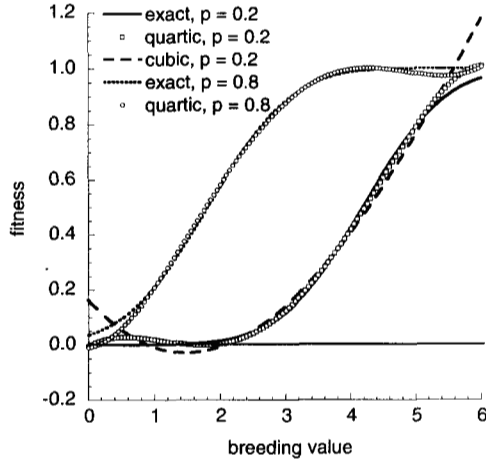
FIGURE 4.—Comparison of the third-order (dashed line) and fourth-order approximations (open symbols) obtained from (32) to the exact truncation fitness function (33) with the fraction selected, p, equal to 0.2 and 0.8 with $V_G = V_E = 1$.
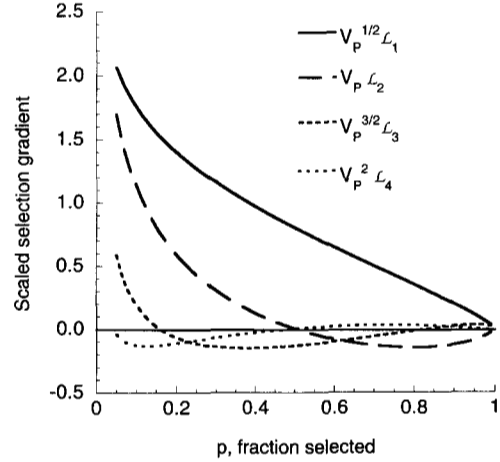

FIGURE 5.—The dimensionless selection gradients, $V_p^{i/2}$ for $i = 1$–4, obtained from (32) and (34) with p ranging from 0.05 to 0.99.

1989, Ch. 11). Applying Equation 26 under the assumption that the distribution of breeding values is Gaussian, we obtain

$$\mathcal{L}_i = \frac{(-1)^{i-1}I(p)}{i!\,\sigma_E^{i-1}\,\sigma_P}\,E[H_{i-1}(hX + \hat{\imath})] \tag{34a}$$
$$\text{for} \quad i = 1, 2, \ldots,$$

where E denotes expectation with respect to a standardized normal density, H_i denotes the ith Hermite polynomial, $h = \sigma_G/\sigma_P$ is the square root of the heritability, and $\hat{\imath} = -t^*\sigma_E/\sigma_P$. The first four $\mathcal{L}_i$ are

$$\mathcal{L}_1 = \frac{I(p)}{\sigma_P}, \qquad \mathcal{L}_2 = \frac{I(p)t^*}{2V_p}, \qquad \mathcal{L}_3 = \frac{I(p)[(t^*)^2 - 1]}{6V_P^{3/2}},$$

and

$$\mathcal{L}_4 = \frac{I(p)t^*[(t^*)^2 - 3]}{24V_P^2}. \tag{34b}$$

$\mathcal{L}_1$ and $\mathcal{L}_2$ are identical to the "parametric" values obtained by differentiating with respect to the parameters of the Gaussian density. The fitness gradients (34b) are identical to those obtained from the polynomial approximations (32) under the assumption

that the distribution of breeding values is Gaussian. [In general, if the initial distribution is Gaussian, both (24) and the polynomial scheme (32) lead to the general expressions provided below for the $\mathcal{L}_i$, see (44) and (48).]

Figure 4 displays the close fit of the fourth-order approximations obtained from (32) to the exact truncation fitness function (33) with $p = 0.2$ and 0.8 and $V_G = V_E = 1$. The polynomial approximations are multiplied by a constant to produce the correct $\bar{W}$. The quartic approximation is very accurate for all breeding values within three standard deviations of the mean, but the cubic approximation is much cruder. The dimensionless selection gradients, $V_P^{i/2}\,\mathcal{L}_i$ for $i = 1$–4, obtained from (32) and (34), are displayed in Figure 5 with p ranging from 0.05 to 0.99. Note that the third- and fourth-order gradients are much smaller than the first two.

Exact recursions for the cumulants: Applying the selection and recombination machinery to an additive polygenic character, we find that the complete recursions up to $\mathcal{L}_4$ (i.e., assuming that $\mathcal{L}_i = 0$ for $i \geq 5$) for the means (κ_i), covariances (κ_{ij}), and third-order cumulants (κ_{ijk}) across loci are

$$\Delta\kappa_i = \kappa_{i\cdot}\,\mathcal{L}_1 + \kappa_{i\cdot\cdot}\,\mathcal{L}_2 + \kappa_{i\cdots}\,\mathcal{L}_3 + \kappa_{i\cdots\cdot}\,\mathcal{L}_4, \tag{35}$$

$$\Delta\kappa_{ij} = \tilde{\Delta}\kappa_{ij} - \Delta\kappa_i\Delta\kappa_j, \tag{36}$$

and

$$\Delta\kappa_{ijk} = \tilde{\Delta}\kappa_{ijk} - [(\tilde{\Delta}\kappa_{ij} + r_{ij}\kappa_{ij})\Delta\kappa_k + (\tilde{\Delta}\kappa_{ik} + r_{ik}\kappa_{ik})\Delta\kappa_j + (\tilde{\Delta}\kappa_{jk} + r_{jk}\kappa_{jk})\Delta\kappa_i] + 2\Delta\kappa_i\Delta\kappa_j\Delta\kappa_k, \tag{37}$$

where

$$\tilde{\Delta}\kappa_{ij} = -r_{ij}\kappa_{ij} + (1 - r_{ij})\kappa_{ij\cdot}\,\mathcal{L}_1 + [2\kappa_{i\cdot}\kappa_{j\cdot} + (1 - r_{ij})\kappa_{ij\cdot\cdot}]\,\mathcal{L}_2 + [3(\kappa_{i\cdot}\kappa_{j\cdot\cdot} + \kappa_{j\cdot}\kappa_{i\cdot\cdot}) + (1 - r_{ij})\kappa_{ij\cdots}]\,\mathcal{L}_3 \tag{38}$$
$$+ [6\kappa_{i\cdot\cdot}\kappa_{j\cdot\cdot} + 4(\kappa_{i\cdot}\kappa_{j\cdots} + \kappa_{j\cdot}\kappa_{i\cdots}) + (1 - r_{ij})\kappa_{ij\cdots\cdot}]\,\mathcal{L}_4,$$

and

$$\tilde{\Delta}\kappa_{ijk} = -r_{ijk}\kappa_{ijk} + (1 - r_{ijk})\kappa_{ijk\cdot}\,\mathcal{L}_1 + \{2(\kappa_{i\cdot}\kappa_{jk\cdot} + \ldots) + 2[(1 - r_{ij})\kappa_{ij\cdot}\kappa_{k\cdot} + \ldots] + (1 - r_{ijk})\kappa_{ijk\cdot\cdot}\}\,\mathcal{L}_2$$
$$+ \{6\kappa_{i\cdot}\kappa_{j\cdot}\kappa_{k\cdot} + 3[(1 - r_{ij})(\kappa_{ij\cdot}\kappa_{k\cdot\cdot} + \kappa_{k\cdot}\kappa_{ij\cdot\cdot}) + \ldots] + (1 - r_{ijk})\kappa_{ijk\cdots}\}\,\mathcal{L}_3 \tag{39}$$
$$+ \{12(\kappa_{i\cdot}\kappa_{j\cdot}\kappa_{k\cdot\cdot} + \ldots) + 6[(1 - r_{ij})\kappa_{ij\cdot\cdot}\kappa_{k\cdot} + \ldots] + 4[(1 - r_{ij})(\kappa_{ij\cdots}\kappa_{k\cdot} + \kappa_{ij\cdot}\kappa_{k\cdots}) + \ldots] + (1 - r_{ij})\kappa_{ijk\cdots\cdot}\}\,\mathcal{L}_4.$$

Here $\tilde{\Delta}$ denotes the change in the cumulants that would be observed if selection were so weak that products of the $\mathscr{L}_i$ could be ignored; and $r_N \equiv 1 - r_{N\varnothing}$ denotes the frequency with which recombination disrupts the loci in the set N (*i.e.*, the fraction of gametes that are not parental haplotypes at these loci). In each equation, we have indicated sums over a particular subscript by "·" (e.g., $\kappa_{i\cdot} = \sum_{j,k} \kappa_{ijk}$). Equation 39 is abbreviated by using "..." to denote the two additional terms in each sum obtained by permuting the subscripts i, j, and k in the first term.

The expression for the fourth-order disequilibrium is more cumbersome. In general,

$$\Delta\kappa_{ijkl} = \tilde{\Delta}\kappa_{ijkl} - [\Delta\kappa_i(\tilde{\Delta}\kappa_{jkl} + r_{jkl}\kappa_{jkl}) + 3 \text{ similar terms}] - [(\tilde{\Delta}\kappa_{ij} + r_{ij}\kappa_{ij})(\tilde{\Delta}\kappa_{kl} + r_{kl}\kappa_{kl}) + 2 \text{ similar terms}]$$

$$+ 2[\Delta\kappa_i\Delta\kappa_j(\tilde{\Delta}\kappa_{kl} + r_{kl}\kappa_{kl}) + 5 \text{ similar terms}] - 6\Delta\kappa_i\Delta\kappa_j\Delta\kappa_k\Delta\kappa_l, \tag{40}$$

where the "similar terms" are generated by permuting the indices. The general form of $\tilde{\Delta}\kappa_{ijkl}$ is unwieldy, so we will restrict attention to exchangeable, unlinked loci (*i.e.*, unlinked loci with equal effects on the character and identical allele frequencies). In this case, for distinct i, j, k and l,

$$\tilde{\Delta}\kappa_{ijkl} = -\tfrac{7}{8}\kappa_{ijkl} + \tfrac{1}{8}\kappa_{ijkl\cdot}\mathscr{L}_1 + (\tfrac{3}{2}\kappa_{ij\cdot}^2 + 2\kappa_{i\cdot}\kappa_{ijk} + \tfrac{1}{8}\kappa_{ijkl\cdot\cdot})\mathscr{L}_2 + (18\kappa_{i\cdot}^2\kappa_{ij\cdot} + \tfrac{9}{2}\kappa_{ij\cdot}\kappa_{ij\cdot} + 3\kappa_{i\cdot\cdot}\kappa_{ijk} + 3\kappa_{i\cdot}\kappa_{ijk\cdot} + \tfrac{1}{8}\kappa_{ijkl\cdot\cdot\cdot})\mathscr{L}_3$$

$$+ (24\kappa_{i\cdot}^4 + 72\kappa_{i\cdot}\kappa_{i\cdot\cdot}\kappa_{ij\cdot} + 36\kappa_{i\cdot}^2\kappa_{ij\cdot\cdot} + \tfrac{9}{2}\kappa_{ij\cdot\cdot}^2 + 4\kappa_{i\cdot\cdot\cdot}\kappa_{ijk} + 6\kappa_{ij\cdot}\kappa_{ij\cdot\cdot\cdot} + 6\kappa_{i\cdot\cdot}\kappa_{ijk\cdot\cdot} + 4\kappa_{i\cdot}\kappa_{ijk\cdot\cdot\cdot} + \tfrac{1}{8}\kappa_{ijkl\cdot\cdot\cdot\cdot})\mathscr{L}_4. \tag{41}$$

These recursions were checked by numerical multilocus calculations as described below. If the fitness function is a fourth-order polynomial, they are exact. If the fitness function is more complex, additional terms involving $\mathscr{L}_i$ for $i \geq 5$ appear. Nevertheless, Equations 35–41 will often provide a very accurate approximation for selection response, as demonstrated below for truncation selection.

Departures from normality at the infinitesimal limit: To predict the deviation from normality under strong selection, we add the recursions (35)-(41) across loci to produce recursions for the cumulants of the distribution of breeding values. Under the additive model, departures from normality arise from linkage disequilibrium and from the fact that only a finite number of loci contribute to the trait (BULMER 1980, Ch. 8; BARTON and TURELLI 1989). At linkage equilibrium, the central limit theorem implies that, under suitable restrictions on the relative contributions of individual loci, the distribution of breeding values will become Gaussian as the number of loci approaches infinity. For a finite number of loci, the distribution will not be Gaussian if the distributions of allelic effects at individual loci are non-Gaussian. To disentangle the roles of disequilibria and the distributions of allelic effects, we will first consider the simpler case in which a very large number of loci with comparable effects contribute to the trait. As discussed by TURELLI and BARTON (1990) and derived in greater generality in APPENDIX B, in the infinitesimal limit only terms of the form κ_U in which the set U contains k distinct indices contribute to the kth-order cumulant for $k \geq 3$. Using this and assuming unlinked loci, Equations 35, 36 and 38 imply the following recursions, complete to $\mathscr{L}_4$, for the mean phenotype and the genetic variance:

$$\Delta\bar{Z} = 2\sum_{i=1}^{n} \Delta\kappa_i = 2\kappa_{\cdot\cdot}\mathscr{L}_1 + 2\kappa_{\cdot\cdot\cdot}\mathscr{L}_2 + 2\kappa_{\cdot\cdot\cdot\cdot}\mathscr{L}_3 + 2\kappa_{\cdot\cdot\cdot\cdot\cdot}\mathscr{L}_4 = V_G\mathscr{L}_1 + K_3\mathscr{L}_2 + K_4\mathscr{L}_3 + K_5\mathscr{L}_4, \tag{42}$$

and

$$\Delta V_G = \tfrac{1}{2}(V_{G,LE} - V_G) + \tfrac{1}{2}K_3\mathscr{L}_1 + (V_G^2 + \tfrac{1}{2}K_4)\mathscr{L}_2 + (3V_GK_3 + \tfrac{1}{2}K_5)\mathscr{L}_3 + (3K_3^2 + 4V_GK_4 + \tfrac{1}{2}K_6)\mathscr{L}_4 - \tfrac{1}{2}(\Delta\bar{Z})^2, \tag{43}$$

where K_i denotes the ith cumulant of the distribution of breeding values and $V_{G,LE}$ denotes the "genic variance" $(2\sum_i \kappa_{ii})$ which remains constant.

If the distribution of breeding values is Gaussian, $K_i = 0$ for $i \geq 3$. Applying Equation 24 to the phenotypic fitness function, we see that

$$\mathscr{L}_1 = \frac{\Delta_s\bar{Z}}{V_P} \quad \text{and} \quad \mathscr{L}_2 = \frac{\Delta_s V_P + (\Delta_s\bar{Z})^2}{2V_P^2}. \tag{44}$$

Thus, (42) reduces to the standard selection response equation, (1); and (43) reduces to BULMER's (1971) equation

$$\Delta V_G = \tfrac{1}{2}h^4\Delta_s V_P + \tfrac{1}{2}(V_{G,LE} - V_G). \tag{45}$$

To abbreviate the exact expression, up to $\mathscr{L}_4$, for ΔK_3, let $O(\mathscr{L})$ denote the terms in (43) proportional to $\mathscr{L}_i$ (*i.e.*, omit the first expression, which is independent of $\mathscr{L}$, and the last expression, which involves $\mathscr{L}^2$),
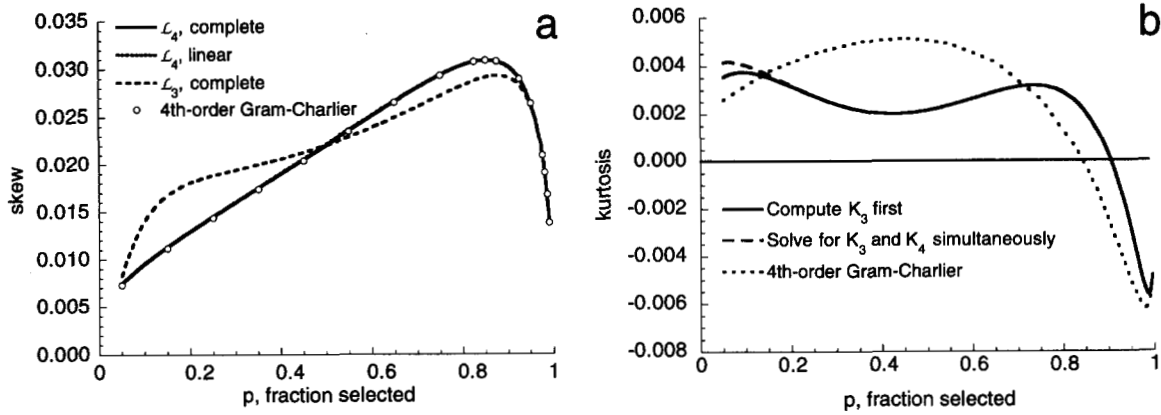
FIGURE 6.—Consequences of continued truncation selection under the infinitesimal model. Graph (a) presents approximations for the equilibrium skew for various p with $h^2 = 0.5$, using (34) for the selection gradients. The "linear" approximation is (49), and it is indistinguishable from the "complete" approximation. Graph (b) presents analogous approximations for kurtosis. See text for details.

then

$$\Delta K_3 = -\tfrac{3}{4}K_3 + \tfrac{1}{4}K_4\mathscr{L}_1 + (\tfrac{3}{2}V_GK_3 + \tfrac{1}{4}K_5)\mathscr{L}_2 + [\tfrac{3}{2}V_G^3 + \tfrac{9}{4}(K_3^2 + V_GK_4) + \tfrac{1}{4}K_6]\mathscr{L}_3$$
$$+ (9V_G^2K_3 + \tfrac{15}{2}K_3K_4 + 3V_GK_5 + \tfrac{1}{4}K_7)\mathscr{L}_4 - \tfrac{3}{2}[O(\mathscr{L}) \text{ terms from } \Delta V_G]\Delta\bar{Z} + \tfrac{1}{2}(\Delta\bar{Z})^3. \tag{46}$$

Similarly, we can abbreviate the exact expression, up to $\mathscr{L}_4$, for ΔK_4 as

$$\Delta K_4 = -\tfrac{7}{8}K_4 + \tfrac{1}{8}K_5\mathscr{L}_1 + (\tfrac{3}{4}K_3^2 + V_GK_4 + \tfrac{1}{8}K_6)\mathscr{L}_2 + (\tfrac{9}{2}V_G^2K_3 + \tfrac{15}{4}K_3K_4 + \tfrac{3}{2}V_GK_5 + \tfrac{1}{8}K_7)\mathscr{L}_3$$
$$+ (3V_G^4 + 18V_GK_3^2 + 9V_G^2K_4 + \tfrac{17}{4}K_4^2 + 6K_3K_5 + 2V_GK_6 + \tfrac{1}{8}K_8)\mathscr{L}_4 \tag{47}$$
$$- 2\Delta\bar{Z}[O(\mathscr{L}) \text{ terms from } \Delta K_3] - \tfrac{3}{2}[O(\mathscr{L}) \text{ terms from } \Delta V_G]^2 + 3(\Delta\bar{Z})^2[O(\mathscr{L}) \text{ terms from } \Delta V_G] - \tfrac{3}{4}(\Delta\bar{Z})^4.$$

A partial consistency check on these recursions is obtained by assuming that the distribution of G is initially Gaussian. In this case, $K_i = 0$ for $i \geq 3$, $\mathscr{L}_1$ and $\mathscr{L}_2$ are given by (44),

$$\mathscr{L}_3 = \frac{\Delta_sK_{3,P} + 3\Delta_sV_P\Delta_s\bar{Z} + (\Delta_s\bar{Z})^3}{6V_P^3} \tag{48a}$$

and

$$\mathscr{L}_4 = \frac{\Delta_sK_{4,P} + 3(\Delta_sV_P)^2 + 4\Delta_sK_{3,P}\Delta_s\bar{Z} + 6\Delta_sV_P(\Delta_s\bar{Z})^2 + (\Delta_s\bar{Z})^4}{24V_P^4}, \tag{48b}$$

where $K_{i,P}$ denotes the ith cumulant of the phenotypic distribution. It follows from (46), (47) and (48) that $\Delta K_3 = h^6\Delta_sK_{3,P}/4$ and $\Delta K_4 = h^8\Delta_sK_{4,P}/8$, as expected.

Truncation selection: Recursions (42), (43), (46) and (47) are generally difficult to apply, because each ΔK_i depends on higher-order cumulants. Thus, no "closed" system of equations can be obtained without assumptions. The simplest generalization of the Gaussian infinitesimal model is to assume that cumulants of order $k \geq 3$ can be adequately approximated by ignoring cumulants of still higher order. This is likely to be accurate only if the cumulants form a decreasing sequence. The consistency of this assumption can be checked by applying it successively to each K_i. We will show that it provides accurate approximations for truncation selection, but fails for strong disruptive selection. According to this scheme, the steady-state value of K_3 under continued selection can be approximated by solving $\Delta K_3 = 0$ with $K_i = 0$ for $i \geq 4$. Figure 6a shows the resulting prediction for K_3 under continued truncation selection with $h^2 = 0.5$, using (34) to approximate the selection gradients. The predicted values are sufficiently small to be very accurately obtained by ignoring terms involving K_3^i for $i \geq 2$. This simple approximation yields

$$K_3 = \frac{2(\mathscr{L}_1^3 - 3\mathscr{L}_1\mathscr{L}_2 + 3\mathscr{L}_3)V_G^3}{3[1 + V_G(\mathscr{L}_1^2 - 2\mathscr{L}_2) + 2V_G^2(\mathscr{L}_2^2 - \mathscr{L}_1^2\mathscr{L}_2 + 3\mathscr{L}_1\mathscr{L}_3 - 6\mathscr{L}_4)]}, \tag{49}$$

which generalizes the weak selection result of TURELLI and BARTON (1990), obtained by ignoring $\mathcal{L}_i^j$ for $j \geq 2$.

Based on our weak-selection approximation, which involved only $\mathcal{L}_3$, we had conjectured that strong selection might produce considerable skew. However, as demonstrated for truncation selection, the additional terms in (49) effectively cancel the contribution from $\mathcal{L}_3$ alone. As shown in Figure 6a, approximation (49) agrees extremely well with the Gram-Charlier statistical prediction. Given that these predictions are based on very different methods and assumptions, we expect that they accurately reflect the behavior of the infinitesimal model. Additional support comes from our finite-locus numerical results below.

A separate problem from determining the number of cumulants needed to accurately approximate the skew is determining the number of selection gradients needed. In the Gaussian infinitesimal model, the variance recursion depends on only the variance, $\mathcal{L}_1$ and $\mathcal{L}_2$. By analogy, one might expect that $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ suffice to predict K_3, at least when it is small. Figure 6a shows that although ignoring $\mathcal{L}_4$ leads to only a small absolute error, it produces a large relative error. Hence, although higher-order disequilibria may be ignored in approximating K_3, its value is significantly affected by $\mathcal{L}_4$. This is also supported by our finite-locus calculations. Figure 6b shows the analogous results for K_4. Essentially identical predictions are obtained by solving sequentially for K_3 from (46) with $K_i = 0$ for $i \geq 4$, then for K_4 from (47) with $K_i = 0$ for $i \geq 5$, versus solving for both K_3 and K_4 simultaneously from (46) and (47). The predicted values of K_4 are all extremely small, but the agreement with the Gram-Charlier predictions is much poorer than for K_3. The significant dependence of K_3 on $\mathcal{L}_4$, illustrated in Figure 6a, suggests that higher order $\mathcal{L}_i$ may be needed to accurately estimate these very small levels of kurtosis. However, as demonstrated below, these very small higher order cumulants are unlikely to affect significantly the response each generation to directional selection.

To quantify the consequences of linkage disequilibrium-induced departures from normality, we will calculate the deviation of the standard Gaussian prediction (1) from the more accurate prediction obtained from (42) by including our predicted steady-state values of skew and kurtosis. The magnitude of the discrepancy is easiest to understand in terms of dimensionless parameters, such as the heritability and standardized cumulants and selection gradients. Let

$$\gamma_{i,G} = \frac{K_i}{\sigma_G^i} \quad \text{and} \quad I_i = \sigma_P^i \mathcal{L}_i, \quad (50)$$

so that $\gamma_{3,G}$ is the skew of the distribution of breeding values, $\gamma_{4,G}$ is the kurtosis; and, under truncation selection, I_1 is the "intensity" of directional selection. In this
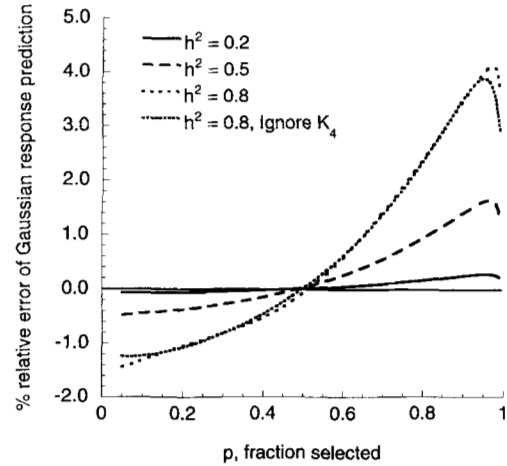

FIGURE 7.—Approximations based on the infinitesimal model for the percent relative deviation of the Gaussian prediction (1) from the exact prediction (51) under long-term truncation selection with various values for h^2 and p. In these calculations, the $\mathcal{L}_i$ are approximated by (34).

notation, Equation 42 becomes

$$\Delta \tilde{Z} = h^2 \sigma_P I_1 \left[1 + h\gamma_{3,G}\frac{I_2}{I_1} + h^2\gamma_{4,G}\frac{I_3}{I_1} + h^3\gamma_{5,G}\frac{I_4}{I_1} \right]. \quad (51)$$

The percent relative deviation of the Gaussian prediction from (51) is displayed in Figure 7 for truncation selection with a wide range of heritabilities and selection intensities. In these calculations, the $\mathcal{L}_i$ are approximated by (34), which shows that the I_i depend on only the fraction of the population selected. The very small deviations from the Gaussian predictions are produced almost entirely by skew. The largest contribution of kurtosis occurs with the highest heritability, but it is noticeable only for extreme selection intensities, i.e., $p < 0.1$ or $p > 0.9$. The qualitative result is that no significant departures from normality are expected from third- or higher-order disequilibria, irrespective of the heritability and selection intensity. For $h^2 \leq 0.8$ and $p < 0.5$, Gaussian theory will underestimate the response to truncation selection by less than 2% per generation, which is likely to be far less than the sampling error of the heritability estimate.

Disruptive selection: Under truncation selection, our results indicate that the small cumulative effect of third-order disequilibria can be accurately approximated by ignoring higher-order disequilibria. We expect this to be true for realistic forms of directional and stabilizing selection. In contrast, disruptive selection can cause significant departures from normality that are much more difficult to predict, because fourth-order cumulants can be very large and their values dependent on still higher-order cumulants. This will be demonstrated with the "double Gaussian" disruptive fitness function (7), denoted $W_{dg}(g)$. As before, we will restrict attention to the (generally unstable) symmetric equilibria with the mean at 0. Because large departures from normality are
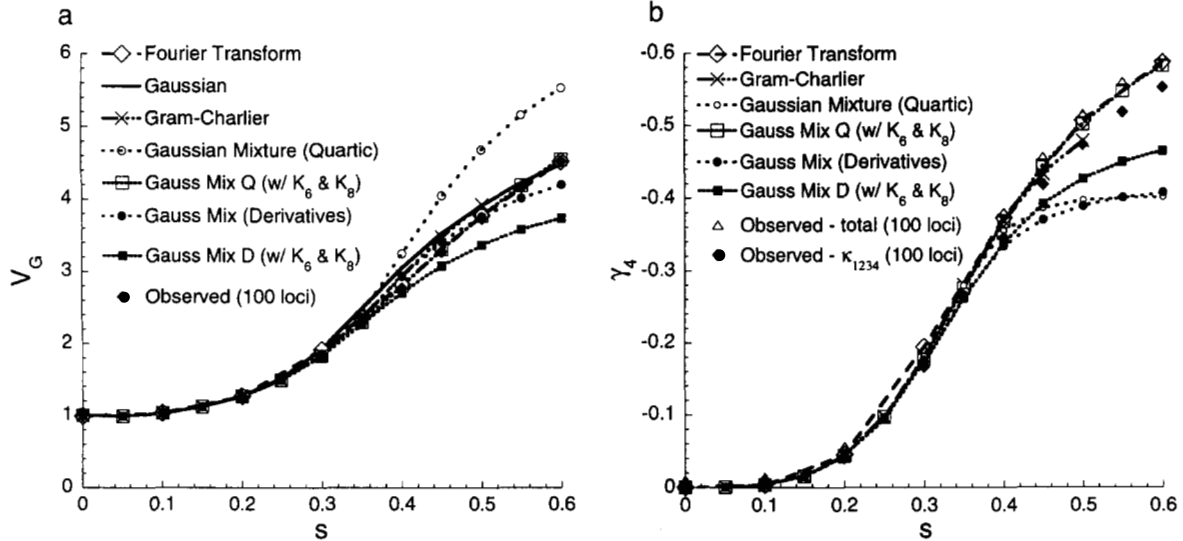
FIGURE 8.—The equilibrium genetic variances (a) and kurtoses (b) under disruptive selection predicted by four alternative genetic approximations based on the "Gaussian mixture" model (52) and three "statistical" approximations (Gaussian, Gram-Charlier and Fourier transform) are compared to numerical results obtained with 100 exchangeable loci. The quartic (Q) approximations calculate the selection gradients from (32), the "derivative" (D) approximations calculate the gradients from (24). The second set of observed values in (b) show only the contribution of fourth-order linkage disequilibria to kurtosis.

expected, we must generalize the Gaussian-based approximations for the $\mathscr{L}_i$ that we applied to truncation selection. To illustrate the effects on non-normality, we will assume that the distribution of breeding values can be approximated by a symmetric mixture of three Gaussians:

$$f_{\text{Mix}}(g) = \tfrac{1}{4}f_{N(-\delta, V)}(g) + \tfrac{1}{2}f_{N(0, V)}(g) + \tfrac{1}{4}f_{N(\delta, V)}(g), \quad (52)$$

where $f_{N(\mu, V)}(g)$ denotes the density of a normal random variable with mean μ and variance V. The parameters δ and V are chosen to produce the appropriate variance and kurtosis, using

$$\text{Var}(G_{\text{Mix}}) = V + \frac{\delta^2}{2} \quad \text{and} \quad K_4(G_{\text{Mix}}) = -\frac{\delta^4}{4}, \quad (53)$$

where G_{Mix} denotes a random variable whose density is (52). The form of (52) is motivated by the strong-selection limit discussed when the Fourier transform method was applied to disruptive selection. More general mixtures than (52), with varying proportions of each Gaussian, produce both positive and negative values of K_4; but we restrict attention to (52) because disruptive selection of form $W_{dg}(g)$ does not produce positive K_4. For instance, if the initial distribution is Gaussian (so that $K_4 = 0$ and the $\mathscr{L}_i$ can computed exactly from Equation 24), $W_{dg}(g)$ produces $K_4 < 0$ in the next generation. The accuracy of approximation (52) is discussed below.

Two different methods were used to approximate the selection gradients with density (52). The first is the quartic approximation (32) based on equating moments of the distribution of breeding values before and after selection. The second is based on identity (24). When the initial distribution is Gaussian, both methods produce identical gradients. They produce slightly dif-

ferent values with the mixture (52). Under our symmetry assumptions, $K_{2i+1} = 0$ for $i = 1, 2, \ldots$ and $\mathscr{L}_1 = \mathscr{L}_3 = \Delta \bar{Z} = 0$, which greatly simplifies Equation 47 for ΔK_4. To illustrate the effect of the higher order cumulants on V_G and K_4, two different approximations for their equilibria will be considered. First, we will set $K_i = 0$ for $i \geq 5$ in (47). This approximation for K_4 is analogous to the simplification we used to predict K_3 under truncation selection. However, assuming $K_i = 0$ for $i \geq 5$ when $K_4 \neq 0$ is unnatural, because (52) implies that

$$K_6(G_{\text{Mix}}) = 4(-K_4)^{3/2} \quad \text{and} \quad K_8(G_{\text{Mix}}) = -34K_4^2. \quad (54)$$

[In general, if for some $i_0 \geq 3$, $K_i = 0$ for all $i \geq i_0$, then all $K_i = 0$ for $i \geq 3$ and the distribution is Gaussian; see STUART and ORD (1987, p. 152).] Hence, our second set of approximations will include the K_6 and K_8 specified by (54) in the recursions for V_G and K_4. The combination of the two ways to approximate the $\mathscr{L}_i$ and the presence or absence of non-zero K_6 and K_8 produces four alternative approximations. Figure 8a shows equilibrium genetic variances predicted by these four approximations and compares them to the Gaussian, Gram-Charlier and Fourier transform approximations presented in Figure 2, as well as numerical multilocus results discussed below. All of the approximations agree for $s \leq 0.3$. However as the intensity of disruptive selection increases, so do the discrepancies between the approximations that use different numbers of cumulants. Thus, our non-Gaussian predictions for both the genetic variance and kurtosis depend significantly on higher-order cumulants. Surprisingly, the Gaussian prediction (and the indistinguishable Fourier transform and Gram-Charlier predictions) remains very close to

the quartic approximation involving K_6 and K_8 based on (52); and these four match most closely our multilocus numerical results.

The apparent accuracy of the Gaussian prediction for the variance is made more surprising by the corresponding values of kurtosis illustrated in Figure 8b. All of the non-Gaussian approximations (and our numerical results) suggest that the distribution of breeding values will have non-trivial kurtosis ($\gamma_4 \leq -0.3$) for strong disruptive selection ($s \geq 0.4$). Nevertheless, the inflation of the variance (the "BULMER effect," e.g., TALLIS 1987) seems to be accurately approximated by the Gaussian-based recursion (45). The basis for this apparent robustness emerges from the more complete recursion (43). For the parameters investigated, the term proportional to $\mathcal{L}_2$ is much larger than the term proportional to $\mathcal{L}_4$. Hence, unless kurtosis becomes very large (e.g., $|\gamma_4| \geq 1$), the primary term driving the BULMER effect is $V_G^2 \mathcal{L}_2$, which appears in the Gaussian analysis and is apparently reasonably approximated by the Gaussian expression for $\mathcal{L}_2$. Similar quantitative results were found for other parameter values. Expression (43) suggests that the robustness of the Gaussian BULMER-effect prediction is likely to extend to other forms of selection that produce significant departures from normality.

Numerical results and approximations for finite numbers of loci: We will present numerical results from multilocus gamete-frequency recursions that check our recursions for the cumulants and explore the accuracy of various approximations for the selection gradients and linkage disequilibria. Two separate computer programs were used. One performs numerical iterations for the genotype frequencies of up to 10 linked, diallelic loci, with arbitrary recombination rates between adjacent loci (assuming no interference) and additive allelic effects $\pm a_i/2$ at the ith locus. This program imposes no constraints on the allele frequencies (see TURELLI and BARTON 1990). For n loci, the state of the population is completely characterized by n allele frequencies (by convention, we give the "+" allele frequency) and all disequilibria up to order n involving sets of distinct loci. Central moments (and cumulants) involving repeated indices can be expressed in terms of these variables by using the reduction formula:

$$C_{U+ii} = C_U C_{ii} - 2m_i C_{U+i}, \qquad (55a)$$

where m_i denotes the average allelic effect at locus i. The second program, described in BARTON (1992), assumes that the loci are "exchangeable," meaning that they are unlinked and have equal allele frequencies and additive effects. The alleles have effects 1 and 0. The symmetry allows for calculations involving up to 100 loci. For n loci, the state of this population is completely characterized by n variables, the frequency of allele "1" (identical for each locus), and a single kth-order disequilib-

rium value for each $k = 2, 3, \ldots, n$. For this model, central moments involving repeated indices can be reduced using

$$C_{U+ii} = p_i q_i C_U - (2p_i - 1) C_{U+i}, \qquad (55b)$$

where p_i denotes the frequency of the "1" allele at locus i (see (A8) in BARTON 1986). To check our recursions for the first four cumulants, we applied quartic selection and then used the numerically observed values of the first eight cumulants in the current generation to predict the values of the first four cumulants in the next generation. No discrepancies larger than expected from round-off error were observed. [Numerical checks on the recursions for the means and covariances are described in TURELLI and BARTON (1990).]

Disruptive selection: We will first discuss the multilocus results for disruptive selection presented in Figure 8. These approximate equilibrium values were obtained by iterating the exchangeable-loci recursions starting from the unstable equilibrium with allele frequency 0.5 at all loci and global linkage equilibrium. The genotypic values were scaled so that this initial state corresponds to a genotypic mean of 0 and genetic variance of $V_{G,LE}$. The values of the second and fourth moments of breeding values closely approach their asymptotic values within 20 generations, during which the allele frequencies remain essentially unchanged. The "equilibrium" values were obtained at generation 25, before the cumulative effects of round-off errors moved the population mean significantly away from the unstable equilibrium at 0. With an infinite number of loci, only fourth-order disequilibrium will contribute to kurtosis of breeding values; whereas for any finite number, there will be contributions from within-locus effects and lower order disequilibria. For this reason, Figure 8b separates the component of kurtosis attributable to fourth-order disequilibrium.

Three basic results emerge from Figure 8 and runs with other sets of parameter values. First, the best multilocus approximation for the 100-locus numerical results is obtained by using the moment-based quartic approximation for the selection gradient and including the estimates for the higher-order cumulants, K_6 and K_8, given by (53). Second, the Gaussian approximation for the genetic variance remains quite accurate even with considerable departures from normality. [BULMER (1980, p. 157) and SØRENSEN and HILL (1983) reported similar agreement between Gaussian predictions for the Bulmer effect and numerical multilocus calculations for "double truncation" disruptive selection. Their numerical results indicate that our conclusion is not an artifact of the exchangeability assumption.] Third, the Gram-Charlier and Fourier transform analyses of the infinitesimal model both approximate the numerical results very accurately. Note, however, that the predictions in Figure 8 are based on the infinitesimal limit; but the
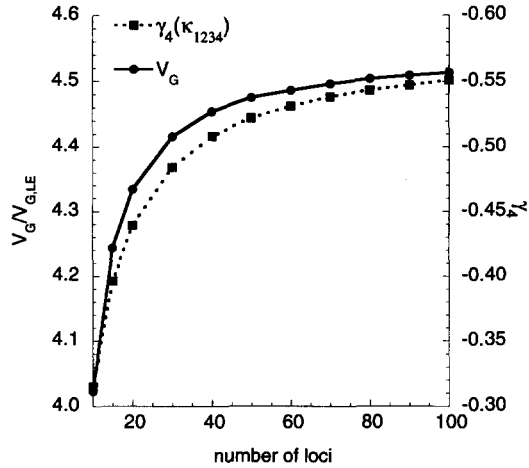
FIGURE 9.—Numerically determined genetic variances and kurtoses under disruptive selection attributable to fourth-order linkage disequilibrium as a function of the number of exchangeable loci. See text for details.

multilocus calculations involve 100 loci. How close is this to the infinitesimal limit? Figure 9 shows how the genetic variance and the kurtosis attributable to fourth-order linkage disequilibrium change with the number of loci. V_G seems to approach an asymptote somewhat faster than γ_4. Hence, it seems likely that the small discrepancies observed between the Fourier transform approximation and the 100-locus calculations would be further reduced by increasing the number of loci. More detailed analyses of finite-locus effects, like those discussed below for truncation selection, seem unwarranted for this model, because it was introduced merely to counter the intuition that additive polygenic inheritance insures approximate normality of breeding values.

Truncation selection: Three issues will be explored: (1) the accuracy of the Gaussian response Equation 1, (2) the accuracy of the Gaussian Bulmer-effect prediction for the consequences of pairwise disequilibrium on genetic variance, and (3) the accuracy of our multilocus predictions for the third-order linkage disequilibrium (and skew) created by selection. Our analysis of the infinitesimal model suggests that linkage disequilibrium will contribute only negligibly to departures from normality. Hence, we expect that unless the number of loci is extremely small, the Gaussian-based predictions for changes in the mean and for the Bulmer effect will be quite accurate. Some numerical results supporting this are presented in Figure 10 a and b. These calculations were performed with the "exchangeable" program, assuming 50 loci, $V_E = 20$, and an initial allele frequency of 0.2 (and hence, an initial heritability of 0.44). The gamete-frequency recursions were iterated for 20 generations, using 20%, 50% or 80% of the population as parents (*i.e.*, $p = 0.2, 0.5, 0.8$). Figure 10a compares the actual change of the mean per generation to the Gaussian prediction (1) based on the mean and additive genetic variance at the beginning of each generation. The

Gaussian response approximation is always quite accurate, the percent relative error of this prediction being less than 4% for all of the values shown. The change in the response per generation reflects changes in allele frequencies caused by selection. The final allele frequencies are 0.91, 0.66 and 0.39, respectively, with 20%, 50% and 80% survival.

Figure 10b compares the observed ratio $V_G/V_{G,LE}$ to the ratio implied by a Gaussian prediction for the BULMER effect. This Gaussian prediction is obtained from Equation 45 by assuming that V_G has reached a "dynamic equilibrium" between the effects of selection and recombination so that $\Delta V_G = 0$ (*cf.* BULMER 1971, 1980, pp. 154–160). BULMER's Gaussian-based approximations yield the following prediction for the genetic variance in terms of the harmonic mean recombination rate, r_h; the current genic variance, $V_{G,LE}$; and the fraction of the population selected as parents, p:

$$V_G = r_h \left(\frac{V_E - V_{G,LE} - \sqrt{(V_E - V_{G,LE})^2 - \frac{2f(p)V_EV_{G,LE}}{r_h}}}{f(p) - 2r_h} \right),$$ (56)

where $f(p) = I(p)[t^* - I(p)]$, with t^* denoting the normalized truncation point and $I(p)$ denoting the intensity of truncation selection, as in Equation 34b. For the exchangeable program, $r_h = 0.5$. Prediction (56) will be referred to as a "generalized quasi-linkage equilibrium" (*GQLE*) approximation for V_G (*cf.* TURELLI and BARTON 1990; BARTON and TURELLI 1991). It rests on the fact that the time-scale for recombination will often be very rapid relative to selection, so that these forces equilibrate rather quickly. Figure 10b shows that within three generations, the values predicted from (56), using the numerically observed values of $V_{G,LE}$, closely approximate the numerical observations. From generations three to 20, the maximum percentage relative error is less than 2%.

The small levels of skew created by third-order disequilibrium are displayed in Figure 10c. These are quantified by $\gamma_3(d) = 2 \sum_{i \neq j \neq k} \kappa_{ijk}/V_G^{3/2}$. The numerically determined values are compared to two different *GQLE* approximations, based on different approximations for the $\mathcal{L}_i$. The "Gaussian" approximation uses (34b). The "quartic" approximation first uses (32) to fit the coefficients of a quartic fitness model under the assumption that the distribution of breeding values is Gaussian, but then uses (29) to take into account the departures from normality caused by within-locus effects (terms like κ_{iii} and κ_{iiii}) and pairwise disequilibria (terms like κ_{iij} and κ_{iijj}). Like the BULMER approximation discussed above, these predictions are obtained by solving $\Delta\kappa_{ijk} = 0$ (see 37 and 39), assuming that all higher-order disequilibria are 0. Although higher-order disequilibria are ignored,
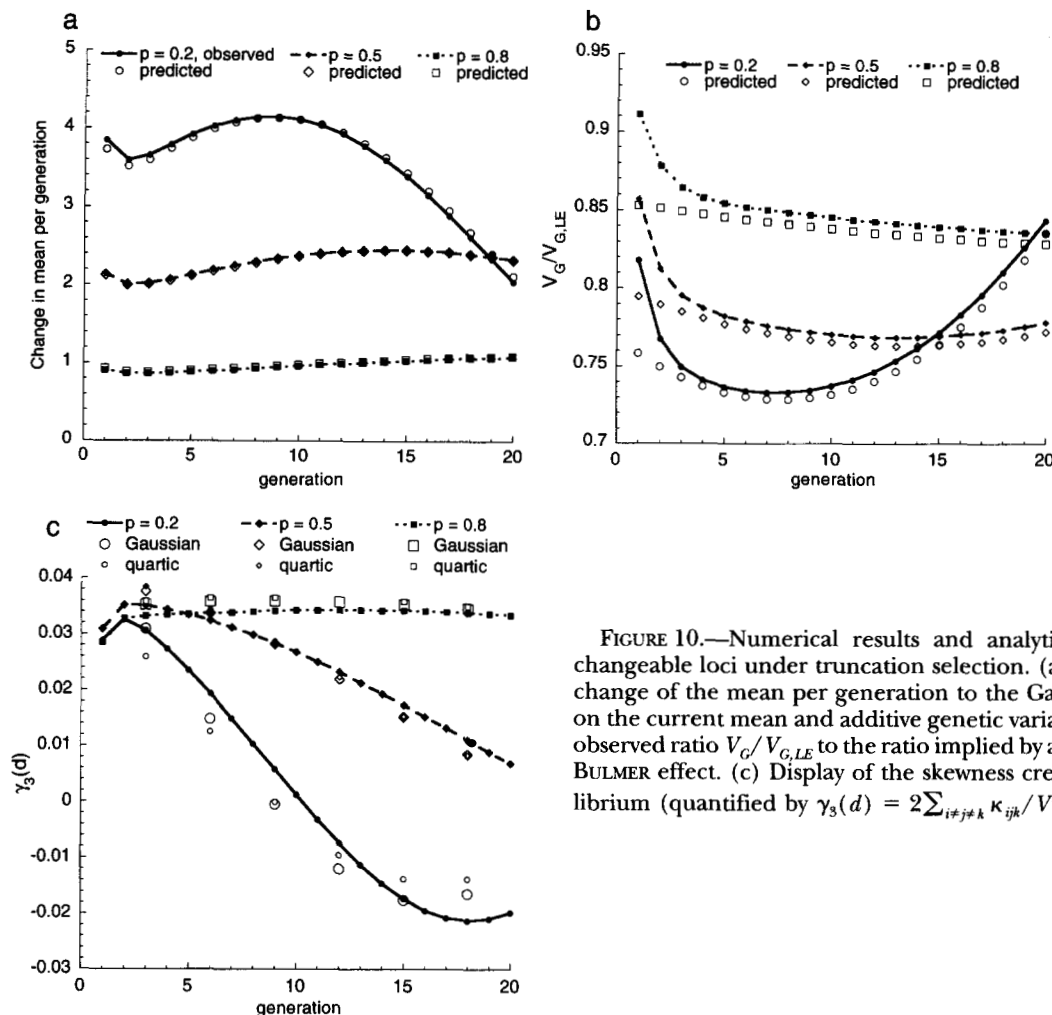
FIGURE 10.—Numerical results and analytical predictions for 50 exchangeable loci under truncation selection. (a) Comparison of the actual change of the mean per generation to the Gaussian prediction (1) based on the current mean and additive genetic variance. (b) Comparison of the observed ratio $V_G/V_{G,LE}$ to the ratio implied by a Gaussian prediction for the BULMER effect. (c) Display of the skewness created by third-order disequilibrium (quantified by $\gamma_3(d) = 2\sum_{i \neq j \neq k} \kappa_{ijk}/V_G^{3/2}$).

the contributions of within-locus effects and disequilibria up to order three are included in higher-order cumulants like $\kappa_{iiijjkl}$ that enter our approximation for $\Delta\kappa_{ijk}$. As expected from our infinitesimal analysis, the observed values of $\gamma_3(d)$ are all extremely small and are adequately approximated by our *GQLE* scheme. However, because the actual values are so small, the percent relative errors can be quite large. Given the apparently negligible biological importance of such terms, more elaborate approximations seem unwarranted. It is worth noting, however, that less elaborate approximations, such as the infinitesimal approximation (49) which ignores all within- and between-locus cumulants of order four or higher, are far less accurate than those presented in Figure 10c.

All of the numerical examples above are based on the "exchangeable" program. To determine the robustness of our conclusions concerning the accuracy of the Gaussian approximations for changes in the mean and the BULMER effect under strong truncation selection, we also performed extensive simulations using a more general multilocus selection program, which allows for linkage, unequal allelic effects and arbitrary allele frequencies.

The results from a "typical" run, with 40% of the population selected as parents, are shown in Figure 11. The closed and open circles compare the actual change in the mean per generation to the Gaussian prediction (1), and the triangles compare the actual ratio $V_G/V_{G,LE}$ to the BULMER prediction (56). These calculations involve 10 linked loci with recombination rate 0.2 between adjacent loci and no interference; thus the harmonic mean recombination rate is $r_h = 0.338$. The allelic effects are $\pm0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.2,$ and 0.2; and the initial frequencies for the "+" alleles are $0.1, 0.2, 0.3, 0.2,$ $0.1, 0.4, 0.6, 0.1, 0.4,$ and 0.4. The calculations were begun with the loci at linkage equilibrium, so that $V_{G,LE} = 3.8984$; we set $V_E = 5.8476$ to produce an initial heritability of 0.4. As shown in Figure 11, the predictions for the change in the mean are always within 2% of the observed values; similarly, after the first four generations, the predicted ratios for $V_G/V_{G,LE}$ are within 6% of the observed values. (The predictions for the first three generations are less accurate, because the "equilibrium" assumption underlying (56) is inappropriate.) Given that much larger statistical errors are expected from estimating variance parameters from actual populations,
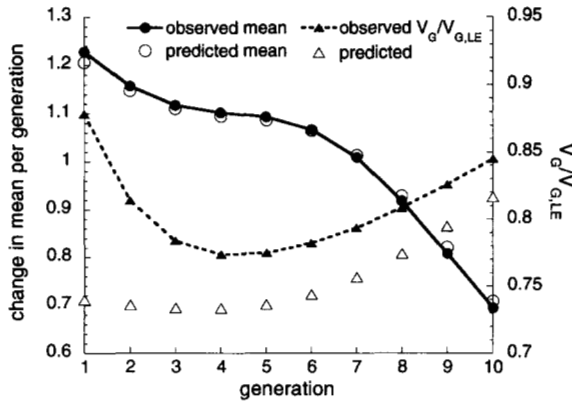
FIGURE 11.—Numerical results and analytical predictions for 10 linked loci with unequal effects under truncation selection (see text for details). The circles compare the actual change of the mean per generation to the Gaussian prediction (1) based on the current mean and additive genetic variance. The triangles compare the observed ratio $V_G/V_{G,LE}$ to the ratio implied by a Gaussian prediction for the BULMER effect.

these errors caused by departures from normality are negligible. Larger departures from the simple BULMER-effect prediction (56) are observed with tighter linkage. However, as noted by BULMER (1980, p. 160) and TURELLI (1984), harmonic mean recombination rates for polygenic traits are unlikely to be much below 0.3 for most higher eukaryotes, with haploid numbers on the order of ten and recombination in both sexes.

DISCUSSION

Until now, almost all analyses of polygenic selection have assumed a normal distribution of breeding values; the few exceptions deal with just two loci, or exponential, quadratic or weak selection (*e.g.*, BÜRGER 1993; GAVRILETS and HASTINGS 1993; ZHIVOTOVSKY and GAVRILETS 1992; NAGYLAKI 1993). Here, we use a more general framework and treat various forms of strong selection, in order to understand why the Gaussian assumption is so successful when applied to additive traits, and in order to develop general methods for multilocus problems. We first presented statistical methods for solving the infinitesimal model, in which the distribution of offspring breeding values is normally distributed around the mean of the parents, with fixed variance. Here, the approximation that the overall distribution is Gaussian gives an accurate prediction for the variance, even when strong disruptive selection generates substantial deviations from normality (Figure 3). We then set out a genetic analysis of selection and recombination. The population is represented by multilocus cumulants κ_U between sets of loci $U = \{ijk\cdots\}$, and selection is described by the relation between mean fitness and the cumulants ($\mathscr{L}_V = \partial \ln(\bar{W})/\partial\kappa_V$). Selection changes the non-central moments by (precisely) $\Delta_s C_U^* = \sum_V (\partial^{|U+V|} \hat{G}^*[\mathbf{0}; \mathbf{0}]/\partial\hat{x}_U\partial\hat{y}_V)\mathscr{L}_V$, where the generating function $\hat{G}^*[\hat{\mathbf{x}}; \hat{\mathbf{y}}] = \exp[K(\hat{\mathbf{x}} + \hat{\mathbf{y}})$

$- K(\hat{\mathbf{y}})] - \exp[K(\hat{\mathbf{x}})]$ and $K(\hat{\mathbf{y}})$ denotes the cumulant generating function for the multivariate distribution of allelic effects (see Equation 19 for the extension to diploid notation and Equations 11 and 12 for the definition of the cumulant generating function). The non-central moments after recombination are calculated as a sum over all the permutations produced by meiosis (Equations 14 and 15 of BARTON and TURELLI 1991). Finally, the cumulants after selection and recombination are calculated from the moments. Although this scheme is applied here in detail only to selection on an additive trait, it is quite general; it allows us to show that even with arbitrary linkage and epistasis, one can define a consistent infinitesimal limit. In this limit, short-term selection response is dominated by infinitesimal allele frequency changes and linkage disequilibria, with the sum of within-locus variances remaining constant (APPENDIX B).

Numerical results show that the Gaussian approximation for the distribution of breeding values gives remarkably accurate predictions for the dynamics of the mean and variance in the infinitesimal limit. With truncation selection, linkage disequilibria of order three and higher never cause much deviation from normality. Disruptive selection can generate substantial four-way disequilibria, and hence kurtosis in the distribution of breeding values; but even then, the Gaussian assumption predicts the variance accurately (Figure 8). Our genetic analysis indicates why only extreme kurtosis is likely to produce large deviations from the normality-based predictions for the dynamics of genetic variance attributable to linkage disequilibrium (the "BULMER effect," see Equations 45 and 43). While our statistical and genetic analyses lead to good approximations for the deviations from normality under truncation selection, they provide no simple intuitive explanation as to why these are so small. For example, our weak selection analysis (TURELLI and BARTON 1990) suggested that the skew should be proportional to the selection gradient favoring skew, $\mathscr{L}_3$, which can become relatively large under intense truncation selection (see Figure 5). However, approximation (49), which is accurate with strong selection, predicts that the skew should be proportional to $(\mathscr{L}_1^3 - 3\mathscr{L}_1\mathscr{L}_2 + 3\mathscr{L}_3)$. With truncation selection at least, these terms nearly cancel, leading to near-normality. Our disruptive selection analyses show that near-normality is not an inevitable consequence of the central limit theorem.

One can construct extreme examples in which the selection gradients favoring skew and kurtosis are large. For example, if $W(x) = 1 + b_4(x^4 - 6V_G x^2)$, x being measured relative to the mean, then $\mathscr{L}_1 = \mathscr{L}_2 = \mathscr{L}_3 = 0$, but $\bar{W}\mathscr{L}_4 = b_4$ (Equations 27–29). Solving the recursion for the cumulants by setting $K_i = 0$ for $i > 8$ gives $K_4 = 0.165V_G^2$ for $b_4 = 0.025/V_G^2$. The Gaussian prediction is $V_G = V_{G,LE}$, since $\mathscr{L}_2 = 0$; but our recursion gives $V_G = 1.039V_{G,LE}$. For stronger selection (but with $b_4 \leq$

$1/(9V_C^2)$, so that W remains positive), no solution exists: starting from a Gaussian distribution gives ever-increasing cumulants. This calculation suggests that *some* forms of selection can generate enough kurtosis that the Gaussian approximation breaks down, even though it does not under truncation or disruptive selection. Whether a wide class of models is well-approximated by the Gaussian assumption, and if so why, remain open questions. However, it seems plausible that most realistic forms of selection will not produce large departures. There is an analogy here with the success of the diffusion approximation for genetic drift, which remains accurate even in populations so small that fluctuations in allele frequency are far from Gaussian (EWENS 1979, Ch. 5).

The success of the Gaussian approximation suggests a general approach to multilocus analysis of additive traits. If the distribution of breeding values is close to normal, then dynamics can be predicted solely from allele frequencies and pairwise disequilibria (the cumulants κ_{ij}), ignoring higher-order disequilibria, such as κ_{ijk}. The mean fitness is now a function of the mean and variance; so higher-order gradients ($\mathcal{L}_U$ for $|U| > 2$) can be neglected, and $\mathcal{L}_i$ and $\mathcal{L}_{ij}$ can be calculated explicitly using Equation 25b or 26. Applying these simplifications, Equations 35–38 become

$$\Delta\kappa_i = (\kappa_{ii} + \kappa_{i\cdot}) \mathcal{L}_1 + (\kappa_{iii} + 2\kappa_{ii}\sum_{j\neq i}\kappa_{ij}) \mathcal{L}_2, \quad (57)$$

and

$$\Delta\kappa_{ij} = -r_{ij}\kappa_{ij} + (1 - r_{ij})(\kappa_{iij} + \kappa_{iij}) \mathcal{L}_1$$
$$+ [2(\kappa_{ii} + \kappa_{i\cdot})(\kappa_{jj} + \kappa_{j\cdot}) \quad (58)$$
$$+ (1 - r_{ij})(2\kappa_{iijj} + \kappa_{iiij} + \kappa_{ijjj})] \mathcal{L}_2 - \Delta\kappa_i\Delta\kappa_j.$$

Here, the sum $\kappa_{i\cdot}$ has been expanded into the part due to within-locus variance, κ_{ii}, and that due to a sum over other loci $\kappa_{i\cdot}$ (so that "$\cdot$" now refers to a sum over distinct loci); higher order cumulants have been truncated to include only pairwise cross-locus effects. Given some assumptions about the within-locus distributions (for example, that only two alleles segregate), Equations 57 and 58 simplify further. Cross-locus disequilibria can, if necessary, be approximated using generalized quasi-linkage equilibrium (*GQLE*, see Equation 56). Even though this eliminates the complications of linkage disequilibria, one must still know the distribution of effects of each locus: this is a fundamental difficulty in understanding quantitative traits, since we are usually ignorant of such genetic details (TURELLI 1984; BARTON and TURELLI 1989).

This approach can be seen as an extension of the infinitesimal model, in which we use the Gaussian assumption to calculate the selection coefficients on individual loci, as well as the response due to pairwise linkage disequilibria. Although the infinitesimal model is elegant,

it is accurate only when many loci are involved (*e.g.*, Figure 10). Recent analyses of response to 10–20 generations of artificial selection have demonstrated significant deviations from the infinitesimal model, probably due to changes in allele frequency at a limited number of loci (*e.g.*, MEYER and HILL 1991; BENIWAL *et al.* 1992). More directly, genes of major effect have been mapped in selected lines (*e.g.*, LÓPEZ and LÓPEZ-FANJUL 1993; SHRIMPTON and ROBERTSON 1988; YOO 1980) and natural populations (MACKAY and LANGLEY 1990). These may contribute significantly to the discrepancies of 20% or greater frequently observed between predicted and realized selection response under artificial selection (SHERIDAN 1988; HILL and CABALLERO 1992), whereas our results imply that skew produced by linkage disequilibrium is not likely to be important. Approximations that combine the effects of linkage disequilibria with those of selection on individual loci are therefore needed; however, our ignorance of the distribution of allelic effects, of pleiotropy and of epistasis remains a major difficulty.

If linkage disequilibrium does not contribute significantly to the frequent departures from the classical response equation, what does? In addition to major gene effects, FALCONER (1989, Ch. 12) summarizes several alternatives in his discussion of asymmetry of selection response. Our treatment is wholly deterministic and ignores the effects of genetic drift, which can be considerable with the small effective population sizes that characterize most selection experiments. Drift can be incorporated into our analyses as outlined in Appendix 1 of TURELLI and BARTON (1990). Our deterministic analyses sought systematic departures from Equation 1, analogous to BULMER's (1971) prediction for changes in the variance, that might be relatively independent of the underlying genetics and hence experimentally detectable. We expect that neither experimentalists nor theoreticians will mourn our "failure." Under artificial selection, departures from Equation 1 will often arise from unavoidable natural selection. This confounding selection may occur on the character itself (ZENG and HILL 1986), on other characters that are genetically correlated with it (LANDE and ARNOLD 1983), or on pleiotropic effects of the alleles that are not apparent in external morphology (see the discussion in KONDRASHOV and TURELLI 1992). Even without epistasis, the connection between standing variation and selection response may be less simple than suggested by Equation 1 and its multivariate generalization.

This report concentrates on quantitative traits. However, understanding linkage disequilibria among many loci is fundamental to other evolutionary issues. Arguments on hitch-hiking and the "genetic load" depend on combining the effects of selection and mutation across the whole genome. Understanding the evolution of sex, recombination, mate choice and so on requires that we know how multilocus selection affects alleles that modify

the breeding system. Numerical or simulation results can be obtained for particular models; however, what is needed is a robust relation between the strength of selection on modifiers, and quantities such as the additive variance in fitness that are (in principle) observable. The methods set out here can be applied to such problems: it may be that the generalized infinitesimal limit (APPENDIX B) is most useful when applied to selection on the whole genome. Several analyses have already successfully exploited the dramatic simplifications that arise from reducing multilocus analyses to a consideration of allele frequencies and pairwise disequilibria. A notable example is CHARLESWORTH's (1990) analysis of mutation-selection balance for deleterious alleles as a force maintaining sexual reproduction.

The observations that quantitative traits are often approximately normally distributed and that their inheritance and selection response can be described adequately by the additive model (with dominance) raise two issues. First, is this evidence that inheritance is in fact approximately additive; or alternatively, is the additive model an adequate approximation for many epistatic models? The latter seems more plausible. One can easily produce models in which epistasis produces arbitrary distributions in the absence of selection; however, these may be biologically unrealistic. On the other hand, at least some biologically motivated models of multilocus epistatic interactions tend to produce mainly additive genetic variance (KEIGHTLEY 1989); so experimental evidence for epistasis must come from an analysis of means rather than variances (cf. COHAN et al. 1989; HARD et al. 1992). The second issue is whether a sexually reproducing population can respond effectively to selection that favors a non-Gaussian distribution, for example, disruptive selection favoring exploitation of two distinct resources. This paper shows that even strong disruptive selection on an additive polygenic trait often produces only a slight deviation from normality. Random mating and recombination constrains the distribution to a near-Gaussian form that produces a lower mean fitness than could be achieved by a more general distribution. This is yet another illustration of the fact that mean fitness is generally not maximized under multilocus selection, because of the breakup of adaptive allelic combinations (KOJIMA and KELLEHER 1961; NAGYLAKI 1993). A sexual population might be able to adapt more efficiently by exploiting non-additive inheritance. This is clearly possible if a major gene is available [for example, in butterfly mimicry, sexual dimorphism, or the bimodal distribution of beak shape in Pyrenestes finches (SMITH 1993)]. However, it may be that no plausible pattern of epistasis can allow much deviation from normality if inheritance is polygenic. Sexual populations must then split into reproductively isolated species if they are to exploit the full range of resources (cf. ROUGHGARDEN 1972). Our framework can be adapted to address such problems.

LITERATURE CITED

BARTON, N. H., 1986 The effects of linkage and density-dependent regulation on gene flow. Heredity 57: 415–426.

BARTON, N. H., 1992 On the spread of new gene combinations in the third phase of Wright's shifting balance. Evolution 46: 551–557.

BARTON, N. H., and M. TURELLI, 1987 Adaptive landscapes, genetic distance and the evolution of quantitative characters. Genet. Res. 49: 157–174.

BARTON, N. H., and M. TURELLI, 1989 Evolutionary quantitative genetics: how little do we know? Annu. Rev. Genet. 23: 337–370.

BARTON, N. H., and M. TURELLI, 1991 Natural and sexual selection on many loci. Genetics 127: 229–255.

BENIWAL, B. K., I. M. HASTINGS, R. THOMPSON and W. G. HILL, 1992 Estimation of changes in genetic parameters in selected lines using REML with the animal model. 1. Lean mass. Heredity 69: 352–360.

BENNETT, J. H., 1954 On the theory of random mating. Ann. Eugenics 18: 311–317.

BULMER, M. G., 1971 The effect of selection on genetic variability. Am. Nat. 105: 201–211.

BULMER, M. G., 1980 The Mathematical Theory of Quantitative Genetics. Clarendon Press, Oxford.

BÜRGER, R., 1991 Moments, cumulants and polygenic dynamics. J. Math. Biol. 30: 199–213.

BÜRGER, R., 1993 Predictions of the dynamics of a polygenic character under directional selection. J. Theor. Biol. 162: 487–513.

CHARLESWORTH, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. 55: 199–221.

COHAN, F. M., A. A. HOFFMANN and T. W. GAYLEY, 1989 A test of the role of epistasis in divergence under uniform selection. Evolution 43: 766–774.

CORNISH, E. A., and R. A. FISHER, 1937 Moments and cumulants in the specification of distributions. Rev. Inst. Int. Statist. 5: 307–320.

EWENS, W. J., 1979 Mathematical Population Genetics. Springer-Verlag, Berlin.

FALCONER, D. S., 1989 Introduction to Quantitative Genetics. Ed. 3. Longman, Harlow, Essex.

FELSENSTEIN, J., 1979 Excursions along the interface between disruptive and stabilizing selection. Genetics 93: 773–795.

FISHER, R. A., 1918 The correlation between relatives under the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52: 399–433.

GAVRILETS, S., and A. HASTINGS, 1993 Maintenance of genetic variability under strong stabilizing selection: a two-locus model. Genetics 134: 377–386.

HARD, J. J., W. E. BRADSHAW and C. M. HOLZAPFEL, 1992 Epistasis and the genetic divergence of photoperiodism between populations of the pitcher-plant mosquito, Wyeomyia smithii. Genetics 131: 389–396.

HILL, W. G., and A. CABALLERO, 1992 Artificial selection experiments. Annu. Rev. Ecol. Syst. 23: 287–310.

KARLIN, S., 1979 Models of multifactorial inheritance. I. Multivariate formulations and basic convergence results. Theor. Popul. Biol. 15: 308–355.

KAUFFMAN, S. A., and S. A. LEVIN, 1987 Towards a general theory of adaptive walks on rugged landscapes. J. Theor. Biol. 128: 11–45.

KEIGHTLEY, P. D., 1989 Models of quantitative variation of flux in metabolic pathways. Genetics 121: 869–876.

KOJIMA, K., and T. M. KELLEHER, 1961 Changes of mean fitness in random mating populations when epistasis and linkage are present. Genetics 46: 527–540.

KONDRASHOV, A. S., and M. TURELLI, 1992 Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. Genetics 132: 603–618.

LANDE, R., 1976 Natural selection and random genetic drift in phenotypic evolution. Evolution **30:** 314–334.

LANDE, R., and S. J. ARNOLD, 1983 The measurement of selection on correlated characters. Evolution **37:** 1210–1226.

LÓPEZ, M. A., and C. LÓPEZ-FANJUL, 1993 Spontaneous mutation for a quantitative trait in *Drosophila melanogaster.* II. Distribution of mutant effects on the trait and fitness. Genet. Res. **61:** 117–126.

MACKAY, T. F. C., and C. H. LANGLEY, 1990 Molecular and phenotypic variation in the *achaete-scute* region of *Drosophila melanogaster.* Nature **348:** 64–66.

MEYER, K., and W. G. HILL, 1991 Mixed model analysis of a selection experiment for food intake in mice. Genet. Res. **57:** 71–81.

MITCHELL-OLDS, T., and R. G. SHAW, 1987 Regression analysis of natural selection: Statistical inference and biological interpretation. Evolution **41:** 1149–61.

NAGYLAKI, T., 1993 The evolution of multilocus systems under weak selection. Genetics **134:** 627–647.

PRESS, W. H., S. A. TEUKOLSKY, W. T. VETTERING and B. P. FLANNERY, 1989 *Numerical Recipes in PASCAL: The Art of Scientific Computing.* Cambridge University Press, Cambridge.

ROUGHGARDEN, J., 1972 The evolution of niche width. Am. Nat. **106:** 683–718.

SHERIDAN, A. K., 1988 Agreement between estimated and realised genetic parameters. Anim. Breed. Abstr. **56:** 877–889.

SHRIMPTON, A. E., and A. ROBERTSON, 1988 The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster.* II. Distribution of third chromosome bristle effects within chromosome sections. Genetics **118:** 445–459.

SLATKIN, M., 1970 Selection and polygenic characters. Proc. Natl. Acad. Sci. USA **66:** 87–93.

SLATKIN, M., 1972 On treating the chromosome as the unit of selection. Genetics **72:** 157–168.

SMITH, T. B., 1993 Disruptive selection and the genetic basis of bill size polymorphism in the African finch *Pyrenestes.* Nature **363:** 618–620.

SØRENSEN, D. A., and W. G. HILL, 1983 Effects of disruptive selection on genetic variance. Theor. Appl. Genet. **65:** 173–180.

STUART, A., and J. K. ORD, 1987 *Kendall's Advanced Theory of Statistics, Vol. 1. Distribution Theory,* Ed. 5. Oxford University Press, New York.

TALLIS, G. M., 1987 Ancestral covariance and the Bulmer effect. Theor. Appl. Genet. **73:** 815–820.

TURELLI, M., 1984 Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. Theor. Popul. Biol. **25:** 138–193.

TURELLI, M., and N. H. BARTON, 1990 Dynamics of polygenic characters under selection. Theor. Popul. Biol. **38:** 1–57.

WOLFRAM, S., 1991 *Mathematica,* Ed. 2. Addison Wesley, Redwood City, Calif.

WRIGHT, S., 1935 Evolution in populations in approximate equilibrium. J. Genet. **30:** 257–266.

YOO, B. H., 1980 Long-term selection for a quantitative character in large replicate populations of *Drosophila melanogaster.* II. Lethals and visibles with large effects. Genet. Res. **35:** 19–31.

ZENG, Z.-B., 1987 Genotypic distribution at the limits to natural and artificial selection. Theor. Popul. Biol. **32:** 90–113.

ZENG, Z.-B., and W. G. HILL, 1986 The selection limit due to the conflict between truncation selection and stabilizing selection with mutation. Genetics **114:** 1313–1328.

ZHIVOTOVSKY, L. A., and S. GAVRILETS, 1992 Quantitative variability and multilocus polymorphism under epistatic selection. Theor. Popul. Biol. **42:** 254–283.

Communicating editor: A. G. CLARK

APPENDIX A: THE RELATION BETWEEN SELECTION GRADIENTS AND POLYNOMIAL COEFFICIENTS

Appendix A of BARTON and TURELLI (1991) set out the relation between the gradients of log mean fitness with respect to the central moments, $L_{U,V}$, and the selection coefficients $a_{U,V}$. However, there was an error in that

Appendix which we correct here; this error does not affect the results in the main part of that paper.

We begin by writing the relative fitness of an individual $W(\mathbf{X}, \mathbf{X}^*)$ as a function of its genotype at loci derived from the maternal gamete ($\mathbf{X}$) and the paternal gamete ($\mathbf{X}^*$). In BARTON and TURELLI (1991), diploid zygotes were assumed to produce equal proportions of male and female haploid gametes. Therefore, the mean contribution of genes at locus i immediately after meiosis must be the same for the two sexes, and is written $m_i = E[\mathbf{X}_i] = E[\mathbf{X}_i^*]$. The moments immediately after meiosis are defined with respect to a hypothetical population of zygotes formed by random union of haploid gametes, and so necessarily $C_{U,V} = C_U C_V$. However, to find the selection gradients with respect to changes in the means and central moments, we must consider perturbations away from this symmetric state. The perturbed means and central moments will be denoted by $\tilde{m}_{\varnothing,i}$, $\tilde{m}_{i,\varnothing}$ and $\tilde{C}_{U,V}$. We will first find the mean fitness as a function of these perturbed variables, and then differentiate to find the selection gradients.

Equation 6 of BARTON and TURELLI (1991) defined the individual fitness as a polynomial function of deviations of maternal and paternal genes, $\zeta_i = X_i - m_i$, $\zeta_i^* = X_i^* - m_i$. Since we are considering perturbations to the mean, we must rewrite this equation in terms of $(\zeta_i + \tilde{m}_{i,\varnothing} - m_i)$ and $(\zeta_i^* + \tilde{m}_{\varnothing,i} - m_i)$, where now $\zeta_i = X_i - \tilde{m}_{i,\varnothing}$, $\zeta_i^* = X_i^* - \tilde{m}_{\varnothing,i}$, and $\tilde{C}_{U,V} = E[\zeta_U \zeta_V^*]$. This was done in Equation A7 of BARTON and TURELLI (1991), which should read

$$\frac{W(\mathbf{x}, \mathbf{x}^*)}{\bar{W}}$$

$$
\begin{aligned}
= 1 &+ \sum_U a_{U,\varnothing}\left[E\left(\prod_{i\in U}(\zeta_i + \tilde{m}_{i,\varnothing} - m_i)\right) - C_U\right] \\
&+ \sum_V a_{\varnothing,V}\left[E\left(\prod_{i\in V}(\zeta_i^* + \tilde{m}_{\varnothing,i} - m_i)\right) - C_V\right] \\
&+ \sum_{U,V} a_{U,V}E\left(\left[\prod_{i\in U}(\zeta_i + \tilde{m}_{i,\varnothing} - m_i) - C_U\right]\right. \\
&\qquad\qquad \left.\times \left[\prod_{i\in V}(\zeta_i^* + \tilde{m}_{\varnothing,i} - m_i) - C_V\right]\right)
\end{aligned}
$$

$$(A1)$$

(the last term was incorrect in the original version).

We now take the expectation of Equation A1 over the perturbed population. Since we will only be taking the first derivative at the point where $m_i = \tilde{m}_{\varnothing,i} = \tilde{m}_{i,\varnothing}$, we can expand the products and discard terms of order $(m_i - \tilde{m}_{\varnothing,i})^2$, $(m_i - \tilde{m}_{i,\varnothing})^2$ and higher:

$$
\begin{aligned}
\frac{\tilde{W}}{\bar{W}} = 1 &+ \sum_U a_{U,\varnothing}[\tilde{C}_{U,\varnothing} - C_U] + \sum_V a_{\varnothing,V}[\tilde{C}_{\varnothing,V} - C_V] \\
&+ \sum_{U,V} a_{U,V}[\tilde{C}_{U,V} - \tilde{C}_{U,\varnothing}C_V - C_U\tilde{C}_{\varnothing,V} + C_U C_V]
\end{aligned}
$$

$$+ \sum_U a_{U,\varnothing} \sum_{i\in U} (\tilde{C}_{U-i}(\tilde{m}_{i,\varnothing} - m_i))$$

$$+ \sum_V a_{\varnothing,V} \sum_{j\in V} (\tilde{C}_{V-j}(\tilde{m}_{\varnothing,j} - m_j)) \tag{A2}$$

$$+ \sum_{U,V} a_{U,V} \sum_{i\in U} (\tilde{C}_{U-i,V}(\tilde{m}_{i,\varnothing} - m_i))$$

$$+ \sum_{U,V} a_{U,V} \sum_{j\in V} (\tilde{C}_{U,V-j}(\tilde{m}_{\varnothing,j} - m_j)).$$

Differentiating and evaluating, we obtain:

$$L_{U,V} = a_{U,V} \tag{A3a}$$

$$\text{for} \quad |U| > 0 \quad \text{and} \quad |V| > 0,$$

$$L_{U,\varnothing} = a_{U,\varnothing} - \sum_V a_{U,V} C_V \tag{A3b}$$

$$\text{for} \quad |U| > 1 \quad \text{and} \quad |V| > 1,$$

$$L_{\varnothing,V} = a_{\varnothing,V} - \sum_U a_{U,V} C_U \tag{A3c}$$

$$\text{for} \quad |U| > 1 \quad \text{and} \quad |V| > 1,$$

$$L_{i,\varnothing} = \sum_{U,V} a_{iU,V}(|U| + 1)C_U C_V, \tag{A3d}$$

and

$$L_{\varnothing,j} = \sum_{U,V} a_{U,jV}(|V| + 1)C_U C_V. \tag{A3e}$$

The sums in Equations A3d, A3e are over all sets U and V, including the empty set $\varnothing$. The error in BARTON and TURELLI (1991) was that the factors $(|U| + 1)$, $(|V| + 1)$ were omitted. These factors arise because there are $(|U| + 1)$ permutations of the coefficients $a_{iU,V}$ that contribute.

There was a typographical error in equation A3 of BARTON and TURELLI (1991), which gives the change in the central moments in terms of the selection coefficients $a_{U,V}$. It should read:

$$\Delta_s C_{S,T} = \sum_U a_{U,\varnothing}(C_{S+U} - C_S C_U)C_T$$

$$+ \sum_V a_{\varnothing,V} C_S(C_{T+V} - C_T C_V)$$

$$+ \sum_{U,V} a_{U,V}(C_{S+U} - C_S C_U)(C_{T+V} - C_T C_V) \tag{A4}$$

$$- \sum_{i\in S} C_T C_{S-i}\Delta_s m_{i,\varnothing} - \sum_{j\in T} C_S C_{T-j}\Delta_s m_{\varnothing,j} + O(a^2).$$

Equation A4 includes the factors C_T and C_S in the fourth and fifth terms that were omitted from the previous version.

APPENDIX B: SCALINGS FOR THE INFINITESIMAL LIMIT

How do the dynamical equations simplify when a very large number of loci contribute to the genetic variance? Here, we describe an infinitesimal limit for an arbitrary model. We make no assumption about the relation between genotype and phenotype, or about the kind of selection that acts. Our results are therefore not restricted to the additive case to which they are applied above: they generalize the standard infinitesimal model (BULMER 1980, Ch. 9) to allow for dominance, epistasis and linkage.

We assume that the effects of each locus become very small, so that the various cumulants (κ) tend to zero as the number of loci, n, increases; the selection gradients ($\mathcal{L}$) remain of order 1. This assumption can be interpreted as follows: the loci affect various traits, which combine to determine fitness. The relation between traits and fitness stays constant, but as the number of loci increases, the effect of each locus on the traits tends to zero. It is nowhere assumed that alleles are *additive* in their effects on fitness or on traits. One could instead suppose that the cumulants stay of order 1, and the selection gradients tend to zero: since all that actually matters is the fitness of each genotype (determined by products of Xs and $\mathcal{L}$s), the way allelic effects are labeled (*i.e.*, the assignment of values to the Xs) is arbitrary. However, this is less easy to interpret in terms of measurable traits.

The derivation hinges on assumptions about how the various κ's scale with n. We suggest a plausible scaling, and show that it is consistent, in that if the cumulants start out of this order or smaller, then they will stay of this order for times small compared with $\sqrt{n}$. We do not prove that if the distribution of genotypes starts with some other scaling (for example, with strong linkage disequilibrium), it will decay to the infinitesimal limit, though this seems likely.

During the period when the infinitesimal limit applies, the dynamics depend only on the genic variances (κ_{ii}) and the cumulants across distinct loci (κ_{ij}, κ_{ijk}, ... with $i \neq j \neq k$ ...). The total genic variance, $V_{G,LE} = 2\sum_i \kappa_{ii}$ stays constant, and so the response of the variances and higher moments to selection is entirely due to linkage disequilibria. (Note that though we refer formally to $2 \sum_i \kappa_{ii}$ as the "genic variance," this is not generally equal to the observable genetic variance of an nonadditive trait at global linkage equilibrium.) Moreover, in the limit, only selection gradients involving distinct indices contribute. If there is no linkage, and loci are exchangeable in their effects on fitness (*i.e.*, $\mathcal{L}_S = \mathcal{L}_{|S|}$ for all sets S), the equations simplify further, and depend only on the sums of linkage disequilibria over all loci. In this case, fitness is a function of the sum of effects over loci: a scale transformation reduces this fully exchangeable model to the additive case.

The scaling assumptions: Denote a set of repeated indices by $I = \{iiii \ldots\}$ and a set of distinct indices by $\{ijk \ldots\}$. More complex sets will be written as $U = IJK\ldots$, where I is a set of one or more i's, and so on. Letters $I, J, K\ldots$ will be reserved for sets involving only a single index i, j, k, whereas S, T, U, $V \ldots$ denote arbitrary sets.

We assume that within-locus cumulants κ_I are of order $n^{-|I|/2}$: thus, κ_{ii} is $O(n^{-1})$ and κ_{iii} is $O(n^{-3/2})$. For an additive trait, this must be true for the κ_{ii} if the genic variance $V_{G,LE} = 2 \sum_i \kappa_{ii}$ is to stay of order 1. It will be true of higher within-locus cumulants if the distribution of allelic effects stays of the same form, but has width that decreases with $1/\sqrt{n}$. This can be seen most easily with biallelic loci, with allelic effects that scale as $1/\sqrt{n}$. The exception to this rule is κ_i, since this represents the within-generation *change* in mean relative to the zygote population, and is $O(n^{-1})$. We show below that in general, $\Delta_s\kappa_I$ is $O(n^{-(|I|+1)/2})$.

We assume that cumulants, κ_U, across sets with distinct (non-repeated) indices, $U = \{ijk \ldots\}$, scale as $n^{-|U|}$. This is plausible, because there are $\approx n^{|U|}$ such coefficients (*i.e.*, $n(n - 1)(n - 2) \ldots (n - |U| - 1) \approx n^{|U|}$ for large n): it is certainly true in the additive case if all the linkage disequilibria are of the same order in n. It is not obvious how to scale more complicated quantities. We assume that the cumulant $\kappa_{i+U} \equiv \kappa_{IU}$ is of order $n^{-[|U|+(|I|+1)/2]}$, where there are $|U|$ distinct indices in the set U. Similarly, κ_{IJU} is of order $n^{-[|U|+(|I|+1)/2+(|J|+1)/2]}$. Thus, each additional repeated index divides κ by $\sqrt{n}$. In general, κ_U is $O(n^{-(\#\text{indices}+\#\text{loci})/2})$ with more than one locus in U, or with a single index, κ_i; κ_I is $O(n^{-(\#\text{indices})/2})$ with one repeated index in I. We show below that if a population starts in linkage equilibrium, selection will generate linkage disequilibria ($\kappa_{IJK}\ldots$) that satisfy these scaling relations for the first and subsequent generations.

In finding the effects of selection and recombination, we will be dealing with polynomial functions of the cumulants, in which each term is a product that contains every element of a given set exactly once. If the set contains only one index, i, then each cumulant in the product, $\kappa_{I'}$, is $O(n^{-|I'|/2})$, and so each term is $O(n^{-|I|/2})$. Next, consider sets $U+I$ in which only one index, i, is repeated. Each cumulant κ_S is of order $n^{-(\#\text{loci} + \#\text{indices})/2}$, unless it consists solely of repeated indices from the set I, in which case it is $O(n^{-(\#\text{indices}/2)})$. Thus, cumulants $\kappa_{I'}$, involving only repeated elements of I are $O(n^{-(|I'|/2)})$ (or $O(n^{-1})$ if $|I'| = 1$), cumulants $\kappa_{U'}$ involving only elements of U are $O(n^{-|U'|})$, and "mixed" cumulants $\kappa_{I'U'}$ involving both are $O(n^{-(|U'|+|I'|/2+1/2)})$. Since in every term of the polynomial, the number of loci and the number of indices sum to $|U| + 1$ and $|U|+|I|$ respectively, each term is of order $n^{-(|U|+|I|/2+m/2)}$, where m is the number of "mixed" cumulants, plus the number of κ_i. Applying the same argument to several sets of repeated indices, we find that each term has the same

form, $O(n^{-(|U|+|I|/2+|J|/2+\ldots+m/2)})$. Now, m is a sum over cumulants contributing to the product, with cumulants containing only elements from U, or only elements from one of I or $J \ldots$ contributing 0; cumulants containing a single element contributing 1; cumulants containing elements from U and one of $I, J, \ldots$ contributing 1; and generally, cumulants containing elements from more than one locus, and from k of the $I, J, \ldots$ contributing k. We will usually only be concerned with the leading terms, where $m = 0$, which will include only "pure" cumulants with two or more indices drawn from either the set of distinct indices U, or from only one of the repeated sets (I or $J \ldots$).

Through most of this appendix, we have not referred explicitly to the distinct sets of loci drawn from the two parental gametes, **x** and **x***. As in the rest of the paper, extension to the full diploid case is trivial, using the convention that indices are considered to be repeated only if they are derived from the same gamete. For example, $\kappa_{i,i}$ is $O(n^{-2})$, whereas $\kappa_{ii,\varnothing}$ is $O(n^{-1})$.

Clearly, there are models that violate these scaling assumptions, for example, where one or a few major loci combine with a large number of modifiers of small effect, so that some of the allelic effects are $O(1)$ instead of $O(n^{-1/2})$. It would be interesting to investigate models with strong epistasis, for example, where fitnesses are assigned to genotypes at random from some distribution (*cf.* KAUFFMAN and LEVIN 1987). Against a fixed genetic background, an allele will then have a large effect-of the same order as the range of the distribution. However, the *average* effect of a substitution-averaged over all genetic backgrounds, as is appropriate near linkage equilibrium-will be small, and of order $1/\sqrt{n}$. For this model, the selection gradients are of order 1 if within-locus cumulants scale as $\kappa_I = O(n^{-|I|/2})$, and the population is in linkage equilibrium. However, we have not been able to demonstrate that the selection gradients are still $O(1)$ in the presence of weak linkage disequilibria, $\kappa_U = O(n^{-|U|})$.

The response to selection: The effect of selection on the non-central moments is derived from the generating function G^* (Equations 16 and 17). In calculating the change in the cumulants, it is simpler to work with the analogous matrix $\tilde{G}$ appropriate to describe the leading-order effects of selection on cumulants:

$$\tilde{G}[\hat{\mathbf{x}}; \hat{\mathbf{y}}] = \exp[K(\hat{\mathbf{x}} + \hat{\mathbf{y}}) - K(\hat{\mathbf{y}}) - K(\hat{\mathbf{x}})] - 1, \quad \text{(B1a)}$$

which gives

$$\tilde{\Delta}_s\kappa_U = \sum_V G[U;V]\mathscr{L}_V, \quad \text{(B1b)}$$

with

$$G[U; V] = \frac{\partial^{|U+V|}\tilde{G}[\mathbf{0}; \mathbf{0}]}{\partial\hat{\mathbf{x}}_U\partial\hat{\mathbf{y}}_V} \quad \text{and} \quad \mathscr{L}_V = \frac{\partial \ln(\bar{W})}{\partial\kappa_V}. \quad \text{(B1c)}$$

Equation B1b gives the change in the cumulants only to first order in $\mathcal{L}_V$ (cf. (2.5) of TURELLI and BARTON 1990). However, we show below that the exact change, $\Delta_s\kappa_U$, is of the same order as the first-order change, $\tilde{\Delta}_s\kappa_U$.

$G[U; V]$ is a polynomial in κ, each term of which contains every element of UV exactly once. (This follows from a dimensional argument, by assigning distinct units to each element of the vectors $\hat{x}, \hat{y}$). $G[U; V]$ has the convenient property that every cumulant in $G[U; V]$ includes elements from U and also from V. To see this, consider $\partial^{|U|}\hat{G}[0, \hat{y}]/\partial\hat{x}_U$. Since the exponential factor reduces to 1 when $\hat{x}$ is set to zero, this has the form of a polynomial in $\partial^{|U'|}[K(\hat{y}) - K(0)]/\partial\hat{x}_U$, where U' is a subset of U. Unless this is further differentiated with respect to $\hat{y}_V$, giving terms of the form $\kappa_{U'+V}$, it will reduce to zero when $\hat{y}$ is set to zero. Hence, every cumulant in $G[U;V]$ must include indices from both U and V.

Consider first the change in within-locus cumulants, κ_I. This is the sum of terms like $G[I; SI'M]$ where S is a set of distinct loci not containing i, I' is a set of indices $\{iii\ldots\}$, and M is a set of b repeated indices, $\{JK\ldots\}$, not containing i or the elements of S. We can divide the whole expression into a sum over classes of terms, each class being defined by $|S|, |I'|, |J|, |K|\ldots$. Each class involves a sum over $O(n^{(|S|+b)})$ terms (corresponding to the number of ways to choose the indices), each of order $n^{-(|S|+|I+I'|/2+|M|/2+m/2)}$ (with m determined by the composition of the sets, as discussed above); the net contribution is therefore $O(n^{-(|I+I'|/2+|M|/2+m/2-b)})$. The leading terms would have no "mixed" cumulants ($m = 0$). However, the matrix $G[I; SI'M]$ contains no terms with cumulants involving only the set M, since these are drawn from the second set alone and would reduce to zero as noted above. Similarly, it contains no terms involving only the set S. Terms with $m = 0$ must therefore have $M = \emptyset$, and so the class of leading order is $SI'M = \{i\}$, giving $G[I; i]\mathcal{L}_i = \kappa_{Ii}\mathcal{L}_i$, which is $O(n^{-(|I|+1)/2})$. The leading terms with $m = 1$ have M, $I' = \emptyset$; the leading component is $G[I; S]\mathcal{L}_S$, and is also $O(n^{-(|I|+1)/2})$. Since this component contains only one mixed term, and yet contains no cumulants with elements from only I or only S, it equals κ_{I+S}. These terms can be written explicitly, giving

$$\tilde{\Delta}_s\kappa_I = \kappa_{I+i}\mathcal{L}_{i,\emptyset} + \sum_S \kappa_{IS}\mathcal{L}_{S,\emptyset} + O(n^{-(|I|/2+1)}), \quad (B2)$$

where the sum is over sets of distinct loci S that do not contain i. We have written the selection gradients here in the explicit diploid form. Since $\tilde{\Delta}_s\kappa_I$ is $O(n^{-(|I|+1)/2})$, and κ_I is $O(n^{-|I|/2})$, this shows that the distribution of effects at each locus changes slowly, over a time scale of $\sqrt{n}$ generations for each cumulant.

This argument identifies the classes of selection gradient that make the leading contributions to the selection response. Provided that selection acts only through gradients of low order, the net response will be of the

same order as the leading class. However, if selection acts through epistatic interactions of very high order $\{\mathcal{L}_S,$ with $|S| \approx n$), the sum over very many classes of terms might lead to a significant cumulative effect. This is a complicated issue that deserves further attention. However, note that with two alleles per locus, there is only one gradient of order n ($\mathcal{L}_{ijk\ldots}$), suggesting that at least in this case, classes of very high order not included in Equation B2 will not have a substantial cumulative effect.

Next, consider the response of the cross-locus cumulants to selection, $\tilde{\Delta}_s\kappa_U$, where U is a set of distinct loci. This is the sum of terms like $G[TQ; SQ'M]\mathcal{L}_{SQ'M}$, where $TQ = U$, Q is a set of distinct loci $\{i, j, k\ldots\}$, S is a set of distinct loci not overlapping with Q, Q' is a set of (possibly repeated) indices $\{IJK\ldots\}$ involving the loci in Q, and M is a set of b repeated indices $\{AB\ldots\}$, not overlapping with the other sets. Each such class includes $n^{(|S|+b)}$ terms, each of order $n^{-(|S|+|T|+|Q+Q'|/2+|M|/2+m/2)}$; the net contribution is therefore $O(n^{-(|T|+|Q+Q'|/2+|M|/2+m/2-b)})$. First, we find the leading terms with no "mixed" cumulants ($m = 0$). $G[U;V]$ contains no terms with cumulants involving only the set M, since these are drawn from V alone. Terms with $m = 0$ must therefore have $M = \emptyset$. Thus, the leading component is $G[TQ; SQ]\mathcal{L}_{SQ}$, which is $O(n^{-(|T|+|Q|)}) = O(n^{-|U|})$, the same order as $\kappa_{TQ} = \kappa_U$, as required. Terms with $m \geq 1$ are negligible.

Elements of the matrix G that connect overlapping sets of loci can be simplified in the infinitesimal limit. First, consider $G[U, V]$, where U and V overlap for the single index i; thus, $U = S + \{i\}$ and $V = T + \{i\}$, with $S \cap T = \emptyset$. From Equation B1,

$$G[S+\{i\}; T + \{i\}]$$

$$= \frac{\partial^{|ST|+2}(\exp[K(\hat{x} + \hat{y}) - K(\hat{y}) - K(\hat{x})])}{\partial\hat{x}_{Si}\partial\hat{y}_{Ti}}$$

$$= \frac{\partial^{|ST|}}{\partial\hat{x}_S\partial\hat{y}_T}\left(\frac{\partial^2 K(\hat{x} + \hat{y})}{\partial\hat{x}_i\partial\hat{y}_i} + \left[\frac{\partial K(\hat{x}+\hat{y})}{\partial\hat{x}_i} - \frac{\partial K(\hat{x})}{\partial\hat{x}_i}\right]\right.$$

$$\left. \times \left[\frac{\partial K(\hat{x} + \hat{y})}{\partial\hat{y}_i} - \frac{\partial K(\hat{y})}{\partial\hat{y}_i}\right]\right)$$

(B3)

$$\times \exp[K(\hat{x} + \hat{y}) - K(\hat{y}) - K(\hat{x})]) \quad \text{at} \quad \hat{x}, \hat{y} = 0.$$

Only terms containing κ_{ii} contribute to leading order, and so only terms in which the first factor $(\partial^2 K(\hat{x} + \hat{y})/\partial\hat{x}_i\partial\hat{y}_i)$ is not further differentiated contribute. Thus, to leading order,

$$G[S + \{i\}; T + \{i\}] \quad (B4a)$$

$$= \left(\frac{\partial^2 K(\hat{x} + \hat{y})}{\partial\hat{x}_i\partial\hat{y}_i}\right)\left(\frac{\partial^{|ST|}\exp[K(\hat{x} + \hat{y}) - K(\hat{y}) - K(\hat{x})]}{\partial\hat{x}_S\partial\hat{y}_T}\right)$$

$$\text{at} \quad \hat{x}, \hat{y} = 0$$

$$= \kappa_{ii}G[S; T] + O(n^{-|S|-|T|-3/2}) \quad \text{for} \quad |S|, |T| > 0.$$

Also,

$$G[i; i] = \kappa_{ii} \quad (B4b)$$

and $$G[i; Ti] = G[Ti; i] = \kappa_{Tii} = O(n^{-|T|-3/2}). \quad (B4c)$$

These relations extend recursively to give $G[U, V]$ where U and V contain distinct indices and overlap at several loci.

We can use these reduction formulas to write $\tilde{\Delta}_s \kappa_U$ explicitly

$$\tilde{\Delta}_s \kappa_U \quad (B5)$$

$$= \sum_S \sum_{Q \subset U} \left(\prod_{i \in Q} \kappa_{ii} \right) G[U - Q; S] \mathcal{L}_{SQ\varnothing} + O(n^{-(|U|+1/2)}),$$

where the sum is over sets of distinct loci S that do not overlap with U, and over all subsets Q of U. Note that $G[\varnothing; \varnothing] = 1$, so that the sum includes the term $(\prod_{i \in U} \kappa_{ii}) \mathcal{L}_{U\varnothing}$, which is the only contribution at linkage equilibrium.

When only two alleles segregate at each locus, cumulants with repeated indices reduce to those with distinct indices, and so the derivation above suffices. With multiple alleles, similar arguments can be used to derive the effects of selection on cumulants involving repeated indices, $\tilde{\Delta}_s \kappa_{IJK}$. These are all of the same order as the κ_{IJK} themselves, showing that our scaling rules are consistent.

Strong selection: The change in the cumulant generating function due to selection, $\Delta_s K(\hat{x})$, is precisely $\log[1 + \Delta_s \hat{f}(\hat{x})/\hat{f}(\hat{x})]$, where $\Delta_s \hat{f}(\hat{x})$ is the change in the moment generating function. The previous section dealt with the first order approximation to this, $\tilde{\Delta}_s K(\hat{x}) = \Delta_s \hat{f}(\hat{x})/\hat{f}(\hat{x})$. An exact expression for $\Delta_s \kappa_U$ can be found by expanding $\log[1 + \Delta_s \hat{f}(\hat{x})/\hat{f}(\hat{x})]$ in a Taylor series, differentiating with respect to the set U, and applying the product rule:

$$\Delta_s \kappa_U = \tilde{\Delta}_s \kappa_U - \frac{1}{2} \sum_{U_1 + U_2 = U} \tilde{\Delta}_s \kappa_{U_1} \tilde{\Delta}_s \kappa_{U_2}$$

$$+ \frac{1}{3} \sum_{U_1 + U_2 + U_3 = U} \tilde{\Delta}_s \kappa_{U_1} \tilde{\Delta}_s \kappa_{U_2} \tilde{\Delta}_s \kappa_{U_3} + \dots \quad (B6)$$

$\tilde{\Delta}_s \kappa_U$ is always of order $n^{-(\#loci + \#indices)/2}$. Since the number of indices represented in the sets $U_1 + U_2 + U_3 \dots$ always sums to $|U|$, and the sum of the number of loci represented in the U_i must be at least equal to the number of loci in U, each of the terms in Equation B6 must be of the same order as $\tilde{\Delta}_s \kappa_U$ or smaller. Allowing for strong selection therefore does not alter the order of the κ_U.

Recombination: Since recombination breaks down associations among loci, it is hard to see how it could generate cross-locus cumulants of higher order in n than is assumed by our scaling rules. This can be demonstrated as follows. From Equation 14 of BARTON and

TURELLI (1991),

$$C_N^{*''} = \sum_{S+T=N} r_{S,T} \left(\frac{C_{S,T}^{*'} + C_{T,S}^{*'}}{2} \right)$$

$$+ (1 - r_N) \left(\frac{C_{N,\varnothing}^{*'} + C_{\varnothing,N}^{*'}}{2} \right). \quad (B7)$$

The non-central moments after recombination, $C_N^{*''}$, are a sum over the non-central moments before recombination (but after selection), $C_N^{*'}$, the sum being taken over all partitions $S + T$ of the set N. The pairs of terms $C_{S,T}^{*'}$, $C_{T,S}^{*'}$ arise because gametes are inherited equally from the two sexes; their frequency may differ if selection acts differently in males and females. (Note that because recombination does not alter the means κ_i, Equation B7 applies whether we use central or non-central moments.) Where the set N contains only distinct indices, every element in the sum is of order $n^{-|N|}$ and so recombination does not change the order of C_N^*. (Note that when U contains distinct indices, C_U^* is of order $n^{-|U|}$). The cumulant κ_N consists of a sum of terms, each containing a product of C's of order $n^{-|N|}$; it therefore also remains of the same order under recombination.

Next suppose N contains repeated indices. Consider $N = UIJK \dots$, where U contains distinct indices not in $IJK \dots$ A dimensional argument shows that the cumulant κ_N after recombination is given by a polynomial in the original cumulants, in which each term contains every index in N exactly once. However, each index may be associated with either of the parental gametes. (For example, the expression for $\kappa_{iijj,\varnothing}$ includes terms like $\kappa_{ii,jj}$ and $\kappa_{i,j}\kappa_{i,jj}$). Each term in $\kappa_{UIJK\dots}$ is of order $n^{-(|U|+|IJK\dots|/2+m/2)}$, where m is contributed by mixed terms. Since $\kappa_{UIJK\dots}$ is itself of order $n^{-(|U|+|IJK\dots|/2+m^*/2)}$, where m^* is the number of sets of repeated indices, recombination could generate terms of higher order only if these had the form $\kappa_U \kappa_I \kappa_J \kappa_K \dots$ or $\kappa_{UI} \kappa_J \kappa_K \dots$, so that $m < m^*$. Such terms do appear in the expressions for $C_{UIJK\dots,\varnothing}^*$; however, since the moments always appear as differences between different partitions of the same set $(e.g., C_{SI,TJK\dots}^* - C_{UIJK\dots,\varnothing}^*)$, these products of pure cumulants always cancel. The leading term in $\kappa_{UIJK\dots}$ is $O(n^{-(|U|+|IJK\dots|/2+m^*/2)})$, as required. Letting κ' denote the cumulants after selection and κ'' the cumulants after selection and recombination, we see, for example, that $\kappa_{iijj,\varnothing}'' = (1 - r_{ij}) \kappa_{iijj,\varnothing}' + r_{ij} \kappa_{ii,jj}' + O(n^{-4})$.

Changes in mean and variance: In this section, we write the contributions of indices from the two gametes explicitly. The change in mean under selection contains a contribution from the within-locus variance κ_{ii}, and from linkage disequilibria with distinct sets of loci, U:

$$\Delta_s \kappa_i = \kappa_{ii} \mathcal{L}_{i,\varnothing} + \sum_U \kappa_{iU} \mathcal{L}_{U,\varnothing} + O(n^{-3/2}). \quad (B8a)$$

Here and below, sums are over sets U of distinct indices,

not including i. The change in genic variance is

$$\Delta_s\kappa_{ii} = \kappa_{iii}\mathscr{L}_{i,\varnothing} + \sum_U \kappa_{iiU}\mathscr{L}_{U,\varnothing} + O(n^{-2}). \quad \text{(B8b)}$$

The change in within-locus variance is $O(n^{-3/2})$, while the within-locus variance is $O(n^{-1})$. Hence, it will remain approximately constant for times $o(\sqrt{n})$. The changes in two-locus cumulants are:

$$\Delta_s\kappa_{ij} = 2\kappa_{ii}\kappa_{jj}\mathscr{L}_{ij,\varnothing} + \sum_U (|U|+1)(\kappa_{ii}\kappa_{jU}\mathscr{L}_{iU,\varnothing} + \kappa_{jj}\kappa_{iU}\mathscr{L}_{jU,\varnothing})$$

$$+ \sum_U \left(\kappa_{ijU} + \sum_{ST=U} \frac{|U|!}{|S|!\,|T|!} \kappa_{iS}\kappa_{jT} \right)\mathscr{L}_{U,\varnothing} \quad \text{(B8c)}$$

$$-\Delta_s\kappa_i\Delta_s\kappa_j + O(n^{-5/2})$$

and

$$\Delta_s\kappa_{i,j} = \kappa_{ii}\kappa_{jj}\mathscr{L}_{i,j} + \sum_U (\kappa_{ii}\kappa_{jU}\mathscr{L}_{i,U} + \kappa_{jj}\kappa_{iU}\mathscr{L}_{j,U})$$

$$+ \sum_U \sum_{S+T=U} \kappa_{iS}\kappa_{jT}\mathscr{L}_{S,T} - \Delta_s\kappa_i\Delta_s\kappa_j + O(n^{-5/2}). \quad \text{(B8d)}$$

The first term in Equation B8c is due to $G[ij; ij]$ and $G[ij; ji]$, the second term to the $(|U|+1)$ permutations

of the form $G[ij; iU]$, and the final term to $G[ij; U]$. The factor $(|U|+1)$ does not appear in Equation B8d because there is only one way of choosing $G[i, j; i, U]$. The factor $(|U|!/|S|!\,|T|!)$ in Equation B8c arises because we are now explicitly distinguishing indices derived from each parental gamete. The indices in S and T both correspond to indices drawn from the same parental gamete; there are $(|U|!/|S|!\,|T|!)$ ways of partitioning the set of indices U from a single gamete into sets S, T. In Equation B8d, this factor does not appear, because the sets S and T correspond to selection $\mathscr{L}_{S,T}$ on some particular set of indices S from one gamete and T from the other. For the same reason, a factor two appears in the first term of Equation B8c, but not in the first term of Equation B8d. After recombination,

$$\Delta\kappa_{ij} = (1 - r_{ij})\Delta_s\kappa_{ij} + r_{ij}\Delta_s\kappa_{i,j} - r_{ij}\kappa_{ij}. \quad \text{(B8e)}$$

When selection is on an additive trait, these expressions simplify substantially. Since $K_{|U|} = \sum_{S+T=U}(|U|!/|S|!\,|T|!)\kappa_{S,T}$, $\mathscr{L}_{S,T} = (|U|!/|S|!\,|T|!)\mathscr{L}_{|U|}$, where $U = S + T$. Using this relation, Equations B8 reduce to Equations 42–43 for selection gradients $\mathscr{L}_1$ to $\mathscr{L}_4$, with no linkage.